Universidad
Carlos III de Madrid

Departamento de Ingeniería Telemática

PROYECTO FIN DE CARRERA

Image Processing and Knowledge-based techniques for Automated Quality Improvement of digital images

Autor: Pablo Casas Muñoz

Tutor: Julio Villena Román

Leganés, Septiembre de 2017

ACKNOWLEDGMENTS

I want to take advantage of this space, to thank first of all my parents, who raised me and knew how to transmit to me the necessary values that make it possible for me to reach this stage in my academic life. Secondly, I want to thank my wife, Irene, for her infinite patience and support during these many years. Third, my special thanks to my director Theodor Letmman and my co-director Julio Villena Román, for their dedication, patience and generosity in guiding me with their knowledge to carry out this thesis.

I also want to thank all those who have somehow helped me with this work, whether it be for a consultation with Word, Matlab, or an algebraic deduction. I thank all those with whom I shared in these long years interminable afternoons of study and practical work, between laughter and theorems. With them I passed, without realizing it, an important stage of my life; bitter and happy moments, anguishes and joys for the results obtained.

Thank you all those who have been encouraging me at any moment. I wish to express my deep gratitude to public education, is a pleasure for me to have reached the end of the road of one of the careers of this prestigious Faculty.

ABSTRACT

With its multidisciplinary nature drawing upon a great variety of areas such as mathematics, computer graphics, computer vision, visual psychophysics, optics, and computer science, the theory of Image Processing needs to be made accessible to practitioners from very diverse backgrounds, from amateur photographers to specialists in communications, medicine, or biology.

The presented thesis focusses on improving quality of images as an automatic process, where we regard aesthetic, physical, perceptual and cognitive approaches (which respectively result in pleasant, identical, realistic, and detailed images) as different levels of the same problem: reproduction accuracy. Main contributions include self-contained fundamental material whose value is likely to remain applicable in a rapidly evolving body of knowledge. A basic strategy followed in its preparation was to provide a seamless integration of well-established theoretical concepts and their implementation using state-of-the-art software tools.

We divide the problem into *i)* low-level algorithmic routines, which resemble early-vision stages, where prior knowledge about natural images is implicitly 'coded' in the algorithm itself; and *ii)* explicit knowledge representation for high-level image processing tasks involving algorithm composition, execution, revision and comparison of candidate solutions.

First, we perform a deep theoretical research of classical as well as emerging paradigms in image quality and its two main dimensions, noise and tone reproduction, with emphasis on information theory and signal processing, but also with inspiration from perceptual sciences and computational photography, thus providing a unifying approach. We focus on edge-preserving smoothing filters as a simple, yet very powerful low-level image processing tool to deal with noise reduction and improved tone reproduction by means of extracting intrinsic components of an image. Finally, we develop, prototype and execute selected low-level image processing operators in MATLAB.

Second, in order to enable end-users to accomplish complex Image Processing tasks while at the same time limiting their cognitive and skill requirements, a system is provided in which expert's knowledge is explicitly stated in the form of rules. Developed with classical knowledge-based techniques and finally implemented in Java, the proposed system allows easy adaptation to specific tasks by exchanging knowledge bases for different areas like computer vision, remote sensing or medical image analysis.

Last, but not least, we are not only interested in solving Image Processing problems, we also want to capture, understand and share the reasoning behind so that others, including non-expert users, can use and build on it.

CONTENTS

Chapter 1

INTRODUCTION

To interact with the environment, we need to constantly acquire, interpret, select and organize the information gathered by our senses. Among them, vision is the most specialized one and, consequently, images[1] play a key role in our lives as *carriers of visual information*. Among the great diversity of image types, we refer here to those produced by capturing light on a two-dimensional light-sensitive medium, i.e. photographic images, as is done by a camera in a very similar fashion to our eyes.

Photographic images are not only used for personal consumption. Virtually every branch of science has sub-disciplines that require collecting image data from the visual universe around us. Indeed, photography has been widely used for scientific and documental purposes since it was invented in 1839, with the development of the daguerreotype. And, so it is, that, in its capability to detain the time, justify, incriminate, change our visual code and appropriate the photographed thing, it frequently "*gives us the sense that we can hold the whole world in our heads – as an anthology of image*" [14].

Despite the fact that the engineering of traditional cameras, lenses and films has reached impressive results, photography has always been limited by the strong constraints of chemistry, optics and analog processes. This led master photographers to create a rich craft and to spend hours to finalize each print, relying on skills that are out of reach of most users. The introduction of digital photography in 1981 together with further advances in technologies underlying not only the capture, but also the transmission, storage and display of images, have created a situation in which the use of images as a means of communicating information has become technologically and economically universal. The above-

[1] An *image* –from Latin *imago*- is an artificial resemblance, either on a two or three-dimensional support, usually of a physical object. More general, the term *image* is commonly used to mean a colorant (used in its most general sense, i.e. ink, wax, dye, silver, phosphors, etc.) arranged in a manner to convey "information", not necessarily on a physical substrate [1].

mentioned constraints have disappeared and the processing possibilities are now endless.

This thesis is about processing digital images. Interest in image processing mainly stems from its many applications. These may be grouped into two principal areas: improvement of pictorial information for human interpretation and processing of image data for storage, transmission, and representation for autonomous machine perception. It encompasses everything from low-level signal processing to high-level image understanding, hence resulting in quite a large and interdisciplinary field drawing upon optics, signal processing, electronics, computer science, pattern recognition, perception science, cognitive science. This thesis keeps that interdisciplinary spirit and looks at image processing from several points of view. It focuses on automatic image quality improvement and uses its formulation as a computational problem in order to link the different viewpoints.

Producing digital images with good brightness, contrast and detail is a strong requirement in several areas like vision, remote sensing, biomedical image analysis, fault detection. Producing visually natural images or transforming the image such as to restore and enhance the visual information within, is a primary requirement for almost all vision and image processing tasks. However, automatic image quality improvement, i.e., a method to yield *better* images without human intervention, is still a notoriously difficult task in image processing [12]. On one hand, while many restoration techniques have been accumulating for more than 30 years, giving rise to a spectacular collection of off-the-shelf algorithms of proved utility in various application areas such as remote sensing or medical engineering, it is still difficult to conceive general enough algorithms applicable to a wide variety of tasks and contexts. On the other hand, without a general standard of image quality that can serve as a design criterion for an image enhancement processor [13], most of the enhancement techniques in existence to date are necessarily empirical and heuristic methods, dependent on the particular type of image[8]. More important, most of these techniques require interactive procedures to obtain satisfactory results, and therefore are not suitable for routine application. Besides requiring the user interaction, many such methods require specification of external parameters, which sometimes are difficult to fine-tune. Finally, the enhancement methods most widely employed treat the spatial information in the image in a global fashion, while in many cases it is necessary to adapt the transformation to the local features within different regions of the image.

Within this context, this research work is concerned with the development of robust algorithms and tools capable of dealing with images coming from very different contexts, as well as a knowledge-based image processing application incorporating Image Processing expertise for aiding inexperienced users

Motivation

This thesis was originally motivated by the work in [2], [3] and [4], where the findings from perceptual, cognitive and computer sciences are used to explore, understand, capture (automate) and eventually share both artist's and Image Processing experts' techniques to make an image more effective.

A familiar problem that arises frequently is the one faced by the person who, equipped with a high grade digital camera (considered to be ideal from now on), aims at capturing an image of the visual scene in front of his eyes, so that it can be stored, transmitted and eventually displayed with the final goal of communicating his visual perception. The experience says that no matter how good the camera, the communications channel or the displaying devices are, often the viewer is disappointed because the reproduction presents noise, blur, lack of detail and bad tone reproduction (colour, bright and contrast). In particular, recorded colour images differ from direct human viewing by the lack of dynamic range compression and colour constancy [4].

We may argue that this situation arises as a direct consequence of the misleading similarities between early stages of HVS (Human Visual System) and imaging devices. In fact, fruitful insight into the problem may be gain by considering three key points that many times are given for granted, which respectively relate to the objective and the subjective nature of the problem.

First, taking an image involves counting photons striking a plane. No matter how good the camera is: there will always be an inherent uncertainty due to light's wave-particle duality, respectively translating in blur and noise distortions present in the image.

Second, while a camera always faithfully snapshots the physical luminance of a scene (whether linearly or nonlinearly), the human visual system regulates and adapts itself to the actual viewing conditions. Indeed, it is widely acknowledged that even low-level visual perception (referred to as *early vision*) is not a mere recording or translation, but an interpretation, which is best understood as an explanation of the physical causes of the retinal image [15]. By interpreting the retinal images and not just storing them, biological systems are prepared to generalize from a particular viewpoint, ambient lighting, size, or distance of a specific image.

Third, notwithstanding the technical advances in image capture systems, image reproduction systems rarely render a precise replica of the original. On one hand, reproductions are generally presented at a different size, resolution, dynamic range and surround environment (and, thus, at a different adaptation level, too). On the other hand, pictures have limitations compared to the real optical flow [4]: they are flat, of finite extent, have a limited field of view, represent the scene from a single point of view, are often static, and they have a limited gamut and contrast. Summarizing in mathematical terms, the

reproduction results at best from applying a linear transformation to both, spatial and range domain of the original scene.

A very important consequence of this is that *the direct recording of the optical flow (i.e. photography) might not result in the most realistic image*; *even physical accuracy alone does not guarantee that the resulting image will have a realistic visual appearance when displayed* [3]. Frequently happens that the preferred reproduction is somewhat different from the original. This is particularly true for photographs.

This motivates the use of ***restoration*** (objectively removing degradation introduced by blur and noise) and ***enhancement*** (matching the subjective visual perception) techniques in order to transform the captured image and improve its quality according to the specific intent of the reproduction, as an artist would do.

While digital photography offers incredible power and Image Processing has become highly specialized with programs implementing more and more complex functionalities, giving rise to a spectacular collection of off-the-shelf algorithms of proved utility in various application areas such as remote sensing or medical engineering, it is still difficult to conceive general enough algorithms applicable to a wide variety of tasks and contexts. While there is a growing basis of complex mathematical theories and techniques oriented to solve specific tasks, little effort has been done to integrate the following

a) visual perception and image processing (even at low levels),
b) image processing techniques from a theoretical point of view, and
c) image processing algorithms to perform complex tasks

Besides, program modules are most of the times integrated just from a low-level point of view, and no support is provided to the user without enough expertise for digital image processing to solve practical problems such the one here considered [2]. Complex image processing tasks require selecting the appropriate algorithms and setting the correct parameters values according to the contents and characteristics of the given image and, therefore, are often difficult to fine-tune. Moreover, extensive experimental work is required to develop image enhancement techniques, in which algorithm composition, execution and control are highly based on empirical or heuristic knowledge. As a result, routine application, when feasible, is rather limited.

Where algorithmic solutions either do not exist or are too complex, artificial intelligence methods and tools can advantageous be introduced. They allow an explicit representation of the problem as well as an operative modeling of the Image Processing expertise, thereby providing self-configuration capabilities to adapt the system behavior in accordance with the problem specifications. Such a system could be successfully used to support non-specialists in Image Processing, enabling them to both develop their own Image Processing applications while limiting their cognitive and skill requirements, and exchanging knowledge bases for different application purposes.

1.2 Approach

This work considers image quality improvement from a *communications* perspective, hence with emphasis on *information theory* and *signal processing*, but also with inspiration from *perceptual sciences* and *computational photography*, thus providing a unifying, holistic approach. Up to date, most of the approaches to image quality regard it as an objective distortion problem, i.e., they consider ***image fidelity*** instead of ***perceived quality***. However, little effort has been done to link both, state the problem in terms of *quality* and *images*, defining the former in relation to what are the latter used for and what requirements these uses impose on them [6], [53], i.e., following a functional approach.

Image quality improvement is here considered as an optimization problem: *given an image, can we produce the most relevant picture for a given purpose?* We differentiate a*esthetic*, *physical*, *perceptual* and *cognitive* approaches to image quality, which respectively result in *pleasant*, *identical*, *realistic*, and *detailed* images. We leave the subjective preference out of the scope of this study and concentrate on cognitive and perceptual intents. Both of them require addressing two key problem areas respectively located at opposite end of the imaging pipeline, namely i) reducing unwanted *physical distortions* introduced in image intensities during the image formation processes (i.e., *noise* and *blur* [2]), and ii) displaying such intensities in a meaningful way (i.e., appropriate *tone reproduction*), which may be regarded as removing *perceptual distortions*.

We respectively refer to these tasks as image *restoration* (cognitive aspect) and *enhancement* (perceptual aspect), whose optimization nature requires the design of specific tools for efficient user interaction. There are essentially three strategies to solve this optimization problem: a) the user can solve it, b) the computer can solve it, or c) the solution might involve both user and computer decisions. The general case is mixed: the computer has to take decisions automatically, but the user wants to keep some control and influence the decisions according to the intended use of the image.

This work stablishes also a parallelism between restoration or enhancement and *early* (i.e., low-level) vision tasks, both motivated by multidisciplinary studies in the field of computational vision [1][11] and inspired by the idea of posing *vision as an unconscious inference of the scene* [9][15] to drive research in image processing techniques and image quality models development. From this point of view, an "image processor", like so much the visual system, must exploit the ecology of images, i.e., it must "know" the likelihood of various things in the world, and the likelihood that a given image-property could be caused by one or another world-property. This world-knowledge may be *hard-wired* (i.e., coded) or *learned*, and may manifest itself at various levels of processing.

[2] Noise and blur may be regarded as unwanted physical distortions that respectively relate to *intensity* and *spatial location* uncertainties. As we will see, removing these inevitable leads to a trade-off between them.

This conveniently leads us to consider image quality improvement as an inference problem: *can we find out the right combination of image processing operations that maximize mutual information of visual perceptions of the sender (the person that took the photograph) and the receiver (the person looking at it)*? The hypothesis is that, posed this way, we can substantially improve the performance of even quite simple models and algorithms.

Along this work it is shown the evolution from the idea of 1) modifying the image by *filtering*, through 2) *estimating* the original image based on available information, to finally 3) creating a new image by means of a) *interpreting* the image and extracting the information within (i.e., separating *intrinsic* components such as *reflectance*, from *extrinsic* ones such as noise, blur or *illuminance*), an analysis task we identify with ***perception***), b) transforming these components in accordance to laws of image formation, and c) *rendering* a new image were *relevant information* has been *enhanced* and *unwanted information* has been compensated for or even eliminated (i.e., a synthesis task similar to ***depiction***).

This evolution clearly follows a *bottom-up* approach in the long term. However, within each step, it is in general followed a *top-down* approach. This can be seen as a reminiscence of bottom-up and top-down processes in vision [11]. From the lowest level of image encoding and representation, through the mid-levels of Image Processing operator design, to the highest level of task description and operator chaining.

More specifically:

- We combine the development of low-level image processing techniques to recover intrinsic components that resemble early-vision mechanisms, and the development of a high-level system incorporating expert knowledge about specific intents (e.g., perceptual, cognitive, etc.) to automatically compose algorithms to perform complex image processing tasks, such as image quality improvement.

- We rely on edge-preserving smoothers (robust estimators) in the spatial domain as a fundamental tool for decomposing an image into its *intrinsic* (reflectance) and *extrinsic* (noise and illuminance) components.[3]

- We approach local tone reproduction based on illuminance-reflectance separation, following classical ideas of Retinex and Horn [10], complemented with ideas from recent research on High Dynamic Range (HDR) tone mapping by Durand and Dorsey [5].

[3] Recent studies show considerable statistical regularities in natural images and scene properties that help tame the problems of complexity and ambiguity in ways that can be exploited by biological and artificial visual systems. See Chapter 2 for an introduction on this.

Due to complexity, we leave out of scope both *colour appearance* models [7] (dealing with attributes such as *lightness*, *brightness*, *colourfulness*, *chroma* and *hue*) and *image appearance* models, which extend upon colour appearance models by incorporating findings about spatial vision, in order to also predict attributes such as *sharpness*, *graininess*, *contrast*, and *resolution*.

Remark that we do not attempt to produce perceptually accurate results, but just credible ones delivered by simple, fast and computationally efficient algorithms. This is motivated by both, the lack of accurate absolute luminance data and simplicity.

1.3 Objectives

This work focuses on the foundation of a principled, unified framework, both at the theoretical and practical level, providing a seamless integration of well-established theoretical concepts, analyzing their interrelations with the aim of providing appropriate efficient tools and techniques that will prove useful for everyone entering the field of Image Quality improvement. A complementary objective was to prepare a material that is self-contained and easily readable by individuals with a basic background in digital image processing, mathematical analysis, and computer programming to afterwards start deeper research in many of the fields that for sure will be left opened here. Monographic works or specific research papers hardly never reach such a global map for thinking about where we are and where should we go.

Visual perception is a complex and difficult field of interdisciplinary convergence of sciences such as biology and neurology, psychology, optics, psychophysics, artificial intelligence, knowledge engineering, and computer vision, as well as formal sciences that, like mathematics, provide sophisticated models of knowledge organization. Establishing a global perspective requires many concepts to be mentioned and related, providing a map that gives sense to the intrinsic connection of the several knowledges, and therefore a certain degree of complexity is unavoidable. Portions of this thesis require that the reader has some experience with linear algebra and calculus, e.g., in chapter 4 "*Noise reduction*". In most chapters of the thesis, however, we have always sought clarity and the greatest simplicity in the explanations in order to provide the reader with the basic ideas without using complex math or formal arguments.

1.4 Outline

This thesis is structured in seven self-contained chapters, each one principally organized by image processing problems and not by mathematical concepts, highlighting the different concepts used and especially indicating where they are applied, concluding with a summary section designed to stimulate discussion of the ideas contained in that chapter. The reference list at the end of each chapter constitutes a set of essential readings in the topics of the chapter.

Chapter 2, "*Background*", reviews relevant concepts in image processing, vision science and image modeling to provide a necessarily brief overview of central concepts on these topics and establish the notation for the rest of the dissertation, serving as a reference source, especially for readers from disciplines other than image processing. Inspired by the early human visual system, restoration and enhancement are first viewed as an estimation problem. This unifies both areas of image processing and places them on common ground with research fields such as visual perception, computer vision or information theory.

Chapter 3, "*Image Quality*", describes and analyzes the motivation, general ideas, and specific algorithms underlying the most representative image quality assessment methods available up to now, putting special stress on their interrelation. It reviews image accuracy in terms of *physical* (objective) match, *perceptual* (subjective) match and *functional* (cognitive) match. These respectively result in *identical*, *photorealistic* and *detailed* images, providing a unifying framework.

Chapter 4, "*Edge-preserving Image Smoothing*", elaborates on the edge-preserving image smoothing framework, as a valuable digital darkroom tool for the task of simplification of visual information, with focus on removing unwanted noise introduced in the image formation process. Other applications in computer graphics and image processing include multi-scale tone management, style-transfer or image editing, where it is often paramount to have the edges, or features in general, preserved by the image coarsening process.

Chapter 5, "*Tone Reproduction*", turns to the other end of the imaging pipeline, where the image is consumed, and addresses the issue of displaying such intensities in a meaningful way, hence regarding image quality not in terms of distortion visibility, but in terms of image appearance. It reviews techniques up to date that build on findings from sensory adaptation mechanisms (such as photoreceptor gain control) and cognitive mechanisms (such as perceptual constancy).

Chapter 6, "*Application*", describes the development from scratch of a knowledge-based system for automating complex image processing tasks intended to improve image quality. As such, it covers recognizing what knowledge is being used to solve the problem, categorizing it and determining the best way to represent it.

Chapter 7, "*Conclusions*", summarizes the thesis conclusions, lists its contributions and outlines directions of future work.

1.5 Sources

In selecting from the huge amount of material available (an initial list of more than 200 articles and books), it has been decided that this thesis should target the reader who wishes to *know how* to study IQ: the pages are filled with facts, but the major goal in writing this thesis is to explain how to reach these facts, not the facts themselves (which are actually used only for motivation purposes). In order to include as many facts as are included here, it was necessary to reduce the level of detail. In each case where this was done, the omitted material, while contributed importantly to the theoretical point that was being made, were not essential to the reader's understanding of that point.

The articles, papers, books and any other material selected for this thesis are among the most influential writing on the topics of image processing and visual perception published in the last twenty years. At the end, the selection reflects a need to provide reasonable coverage of the fundamental concepts whose value is likely to remain applicable in a rapidly evolving body of knowledge, and a desire to include articles that are accessible by non-experts. Also included are some research articles and ideas from the last decade to reflect some recent trends in the field. Obviously, many important topics and many influential works have been omitted. It would not be difficult to compile three more volumes of references of equivalent quality and impact. Reviews on implied fields of research are bound to be incomplete and the reader is prompted to the references to fill the numerous gaps that for sure will find.

Concluding this introduction, we hope that readers will enjoy reading this thesis work as much as we have enjoyed its research and writing. We hope also they find materials provided in it timely, stimulating, useful and relevant to their work and studies in the field. We insist that the framework proposed in this thesis is not intended as a rigid set of boxes, but to provide a common framework and raise issues. Although it offers some practical insights, it is intended more as an in-breadth overview and starting point.

REFERENCES

[1] BARROW, Harry G.; TENENBAUM, Jay M. Computational vision. *Proceedings of the IEEE*, 1981, vol. 69, no 5, p. 572-595.

[2] CHRISTENSEN, Henrik I.; CROWLEY, James L. (ed.). *Experimental environments for computer vision and image processing*. World scientific, 1994

[3] DURAND, Frédo. An invitation to discuss computer depiction. *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*. ACM, 2002. p. 111-124.

[4] DURAND, Frédo. Limitations of the medium and pictorial techniques. *Perceptual and Artistic Principles for Effective Computer Depiction, Course Notes for ACM SIGGRAPH 2002*, 2002, p. 27-45.

[5] DURAND, Frédo; DORSEY, Julie. Fast bilateral filtering for the display of high-dynamic-range images. En *ACM transactions on graphics (TOG)*. ACM, 2002. p. 257-266.

[6] ENGELDRUM, P.G. Image Quality Modeling: Where Are We? *IS&T's PICS Conference,* 1999.

[7] FAIRCHILD, Mark D. *Color appearance models*. John Wiley & Sons, 2013.

[8] GONZALEZ, R.C. and WOODS, R.E. *Digital Image Processing*. 2nd ed. Prentice-Hall, 2002.

[9] KNILL, David C.; RICHARDS, Whitman (ed.). *Perception as Bayesian inference*. Cambridge University Press, 1996.

[10] LAND, Edwin H. Recent advances in Retinex theory and some implications for cortical computations: color vision and the natural image. *Proceedings of the National Academy of Sciences*, 1983, vol. 80, no 16, p. 5163-5169.

[11] MARR, David; VISION, A. A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*, 1982, vol. 1, no 2.

[12] MUNTEANU, Cristian; ROSA, Agostinho. Gray-scale image enhancement as an automatic process driven by evolution. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2004, vol. 34, no 2, p. 1292-1298.

[13] PRATT, William K. *Introduction to digital image processing*. CRC Press, 2013

[14] SONTAG, S. *On Photography*. Picador USA, 2001.

[15] VON HELMHOLTZ, Hermann; SOUTHALL, James Powell Cocke. *Treatise on physiological optics*. Courier Corporation, 2005.

[16] WANG, Zhou; BOVIK, Alan C. *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

Chapter 2

BACKGROUND

INTRODUCTION

The ideas in this thesis rely on fundamental principles in image processing. Several authors have pointed out that it is of central importance that an image processing framework must be physically consistent with the nature of the images, and within the context of visual models. Therefore, this chapter reviews relevant concepts in image processing, vision science and image modelling in order to provide a necessarily brief overview of central concepts on these topics and establish the notation for the rest of the dissertation, serving as a reference source, especially for readers from disciplines other than image processing. Background on the more specific topics of image quality and intelligent image processing is provided in the corresponding chapters.

This chapter is organized as follows. Section 2.1, *"Digital Image Processing"* first gives a brief overview of image processing issues, focusing on image restoration. The ill-posedness of the image restoration problem and regularization techniques is then briefly addressed. Section 2.2, "*Vision Models"* introduces basic aspects of early vision from a computational approach based on the Bayesian formulation of visual perception. The connection to regularization techniques of previous section is also established. Section 2.3, "*Overview of image modeling*", presents three fundamental mathematical representations of images. Section 2.4 finally concludes this chapter with a summary of the most important ideas.

2.1 Digital Image Processing

This section provides a brief overview of image processing issues, focusing only on image restoration and enhancement, posed as inverse problems.

In its broad acceptation [22], the notion of processing an image involves the transformation of that image from one form into another, either a new image, abstraction, parametrization or decision. It encompasses everything from low-level signal enhancement to high-level image understanding, hence resulting in quite a large and interdisciplinary field drawing upon optics, signal processing, electronics, computer science, pattern recognition, perception science, cognitive science, and many other related disciplines, a few of which have been gathered in the lower half of Table 2.1. Good popular references in general image processing are the books by Bovik [17], Gonzalez et al. [18], Pratt [20], Russ [21], Jähne B. [19], and the online tutorial in [24].

The first work in image processing dates back to the 1920s, when automated means of image transmission were first used. In the 1950s, computers were starting to be used and, since then, their successive development has conditioned the history of digital image processing, primary concerned with the development of computer algorithms working on digital images [18]. Digital offers over photographic or electrical analog image processing the main advantages of precision and flexibility. Besides, advances in computer technology have significantly reduced its expense and increased its speed.

2.1.1 Levels of Processing

Although boundaries between image processing, image analysis and computer vision are not well established, there is a general agreement on considering low-, mid- and high-level processes [18], [24]. Low-level processes involve primitive operations such as image preprocessing to reduce noise, contrast enhancement, and image sharpening. A low-level process is characterized by the fact that both its inputs and outputs are images. Mid-level processes on images involve tasks such as segmentation, description of features to reduce them to a form suitable for computer processing, and classification (recognition) of individual objects. A mid-level process is characterized by the fact that its inputs generally are images, but its outputs are attributes extracted from those images (e.g., edges, contours, and the identity of individual objects). Finally, higher-level processing involves "making sense" of an ensemble of recognized objects, as in image analysis, and, at the far end, performing the cognitive functions normally associated with human vision.

In addition to the diversity of image types that arise and which can derive from nearly every type of radiation [17], interest in image processing mainly stems from its many applications. These may be grouped into two principal areas: improvement of pictorial information for human interpretation and; processing of image data for storage, transmission, and representation for autonomous machine perception [18].

In general, such image processing problems are solved by a chain of tasks (see Table 2.1), which traditionally consists of the following steps:

1. ***pre-processing and filtering***, resulting in a modified image with the same dimensions as the original image.
2. ***data reduction***: extracting significant components from an image, e.g. edges, texture characteristics or landmarks.
3. ***segmentation***: partitioning an image into regions which are coherent with respect to some criterion.
4. ***object recognition***: determining the position and, possibly, the orientation and scale of specific objects in an image, and classifying these objects).
5. ***image understanding***: obtaining high level -semantic- knowledge of what an image shows).

(These are supported by auxiliary optimization techniques)

	LOW-LEVEL VISION	MID-LEVEL VISION	HIGH-LEVEL V.
ACTIVITY	IMAGE PROCESSING	IMAGE ANALYSIS	IM. UNDERSTANDING
	ENCODING	REPRESENTATION	INTERPRETATION
ABSTRACTION LEVEL	Pixel level	Structure level	Object set level
	Local feature level	Object level	
DATA/INFO.	Proximal stimulus *2D intensity image Available information*	Descriptors: *shading, edges, texture, depth, motion*	Conscious Information
STAGE/STEP	PRE-PROCESSING → DATA REDUCTION → SEGMENTATION → OBJECT RECOGNITION → IMAGE UNDERSTANDING		
AREAS of IMAGE PROCESSING/ TASKS	Restoration & Enhancement: *Denoising deblurring*; *Tone & Detail Management*	*Compression, Feature extraction*; *Texture & Color Segregation, Recognition, Clustering*	*Template matching Feature-based recognition*; *Scene analysis, Object arrangement*
FIELD	*Image Processing*	*Computational & Cognitive Vision, Machine Learning*	
	Imaging; Applied Statistics; Signal Processing; Computer Vision; Artificial Intelligence		
APPLIED SCIENCE	*OPTICAL ENGENIEERING; APPLIED MATHEMATICS; COMPUTER SCIENCE*		
PERCEPTION THEORY	*EWALD HERING* Light adaptation, Local interactions	*GIBSON* GESTALT Laws of Perceptual Organization	*HELMHOLTZ* Unconscious Inference
DISCIPLINE	*Optics; Physiology ; Psychophysics; Cognitive Neuroscience*		
SCIENCE	*PHYSICS; VISUAL ANATOMY; BIOLOGY / NEUROSCIENCE, Systems Neuroscience*		
DEPICTION	*PHOTOGRAPHY POINTILLISM*	*IMPRESSIONISM*; *LINE DRAWING*; *SCHEMA*	*PRIMITIVE ART, CUBISM*

Table 2.1. Visual processing framework. Each step builds upon several applied sciences and may be finally related to an own depiction technique. At the initial, primarily iconic levels (edges, regions, gradients), the processing is data driven, constrained by what is possible to compute directly from the image. At the highest, primarily symbolic levels (surfaces, objects, scenes), the processing is goal driven, dictated by the information required to support the ultimate goals. In between, the order of representations is constrained by what information is available at preceding levels and what is required by succeeding ones, with low-level cues making bottom-up proposals which are validated by high-level models, combining both, bottom-up and top-down processing **[30]**. With each step in the chain the need for using prior (world) knowledge increases. For simple noise reduction, not much knowledge about the contents of the image itself needs to be known, whereas for image understanding it is imperative to limit the domain of images which can be processed.

2.1.1.1 Image processing as low-level or early vision

In this thesis, we deal only with low level image processing operations which do not presuppose any specific type of image content. We will concentrate only on the use local filtering techniques in spatial domain to keep intuitiveness and simplicity. In particular, we will pose image processing as an estimation problem. Early processing techniques include filtering, edge operators, range transforms, computation of surface orientation and optical flow, etc. We concentrate on filtering, a very general notion of transforming the image intensities in some way to enhance or deemphasize certain features. We consider only transforms that leave the image in its original format: a spatial array of gray/colour levels. We cannot, however, afford to examine these techniques in detail here; instead, my intent is to describe a set of techniques that conveys the principal ideas. In this thesis we restrict ourselves to low-level image processing (specifically image restoration and enhancement) in the spatial domain. The following sections address image structure and representation, restoration and enhancement, classical algebraic approach as well as regularization models.

2.1.2 Image Structure and Representation: Notation

Before introducing the issues of image restoration and enhancement, We will introduce some useful notation about image's structure and representation.

An image is generally defined as a real or complex-valued function of two space variables belonging to some support region. Although this support may be continuous, it is commonly sampled on a rectangular grid. The image is then represented as a two-dimensional lattice of p-dimensional vectors (pixels), where $p=1$ in the gray-level case, $p=3$ for colour images, and $p>3$ in the multispectral case.

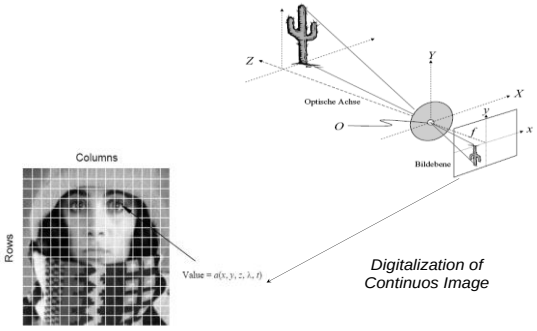


Figure 2.1. Image formation

The space of the lattice is known as the spatial domain (***S***), while the gray level, colour, or spectral information is represented in the range domain (***R***). For both domains, Euclidean metric is assumed. To leave room for a more general case, the range domain is sometimes also referred to as *feature* domain (e.g. for dealing with texture).

Using a discrete formulation, let v_i denote the observed value, sampled at locations sites $x_i=[x_1,x_2]_i^T$. Often a joint domain representation is convenient and vectors $z_i = (x_i^T, v_i^T)^T$ are referred to as generalized pixels' intensities of the (observed) image. This defines a 2D surface embedded in a 3D space for gray-level images, and a 2D surface embedded in a 5D space for colour images [113]. For convenience of notation, we use symbols like x instead of {X=x} in statistics.

2.1.3 Image Improvement: Restoration vs. Enhancement

This thesis deals with several classic problems within the fields of *image restoration* and *image enhancement*. The former deals with processing corrupted or degraded image data in order to reconstruct or recover the uncorrupted image. Removing blur and noise are examples of restoration operations. This is typically performed based on an ***observation model***, which relates the observed degraded image to the desired original image, and possibly a ***regularization model***, which conveys the available *a priori* information about the original image. Thus restoration techniques are oriented toward modeling the degradation and applying the inverse process in order to recover the original image. This approach usually involves formulating a criterion of goodness that will yield an optimal estimate of the desired result.[4]

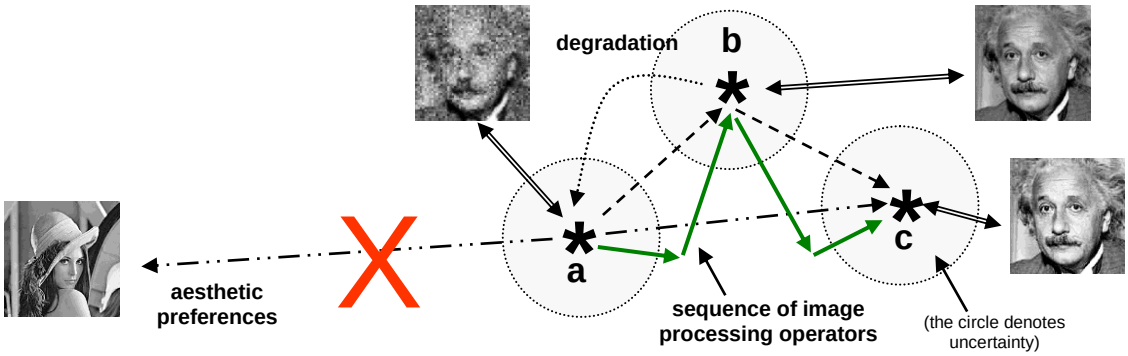


Figure 2.2. Image *Restoration* vs. *Enhancement*: Image quality improvement viewed as a sequence of steps that yield, from the observed degraded image, ***a***, a better image , ***c***. The original unobserved image, ***b***, represents an intermediate stage that breaks down the process into ***restoration*** and ***enhancement***. Notice that in all this, we disregard aesthetic preferences.

Image enhancement, on the other hand, refers to producing visually natural images or transforming the image such as to enhance the visual information within. Tone and detail management (e.g., increasing contrast or revealing details) are typical image enhancement issues. The task of image enhancement is a difficult one considering the fact that there is no general unifying theory of image enhancement at present, because there is no general standard of image quality that can serve as a design criterion for an image enhancement processor [20]. In fact, most of the enhancement techniques in existence to date are empirical or heuristic methods, dependent on the particular type of image [8].

In what follows, we will concentrate in providing introductory ideas about image restoration, only from the point where a degraded digital image is given. Image enhancement will be considered in turn in Chapters 3 and 5. Specifically, we deal with degradation, algebraic and regularization models.

---

[4] Most of the restoration techniques are based on a least squares criterion of optimality. The use of the word *optimal* in this context refers strictly to a mathematical concept, not to optimal response of the human visual system. In fact, the present lack of knowledge about visual perception precludes a general formulation of the image restoration problem that takes into account observer preferences and capabilities [18].

2.1.4 Observation model

We assume that we are dealing with images formed from light using modern electro-optics[5], where the number of electrons counted, N, can be written as $N=N_I+N_{th}+N_{ro}$, where N_I is the number of electrons due to the image, N_{th} the number due to thermal noise, and N_{ro} the number due to read out effects.

$$(N_I)_i = T\int_{\lambda}\int_{\mathbf{x}} B(\mathbf{x},\lambda)S_r(\mathbf{x}-\mathbf{x}_i)q(\lambda)d\mathbf{x}d\lambda$$

where T is the integration time (in seconds), x is a vector representing the continuous coordinates on the sensor plane, $\int_x$ denotes integration over the area of the collection site, x_i are the coordinates of the center of the collection site, and $S_r(x)$ is defined as the response of a collection site that is centered at $x=(0,0)$. $q(\lambda)$ is the device quantum efficiency, defined as the ration (electrons/Joule) of electrons collected per incident light energy for the device as a function of the wavelength λ. The spectral irradiance pattern $B(x,\lambda)$ incident on the sensor is given by

$$B(\mathbf{x},\lambda) = \left[R(\mathbf{x},\lambda)L(\mathbf{x},\lambda) * p(\mathbf{x},\lambda)\right]t(\lambda)$$

where * denotes spatial convolution, $p(x,\lambda)$ is the point spread function of the optics, and $t(\lambda)$ is the spectral transmission of the optics.

Following a discrete formulation, let that u(**x**):=R(x)L(x) denote a function describing the original, unobservable, image and $v_i:=N_x$ denote the observed value, sampled at locations/sites $x_i=[x_1\ x_2]_i^T$. The degradation process is then usually modeled as an operator A that, together with an additive noise term n, operates on the input image $u=u(x)$, to produce the observed degraded image v: $\mathbf{v} = \mathrm{A}(\mathbf{u}) + \mathbf{n}$ *(2.1)*.

While the particular form of the operator A depends upon the particular assumptions that are made about the image formation process, often it can be linearly modeled, leading to the simpler model $\mathbf{v} = \mathbf{A}\cdot\mathbf{u} + \mathbf{n}$ *(2.2)*, where A is now a degradation matrix. In addition, if A is shift-invariant[6] (e.g. A represents a blurring process), the former expression may be formulated as a convolution operation $\mathbf{v} = \mathbf{a} * \mathbf{u} + \mathbf{n}$ *(2.3)*, where a is now the equivalent ***impulse response*** or ***point spread function*** (PSF).

The noise term n typically accounts for errors and uncertainties in the image acquisition process and is assumed to be a zero mean independently and identically distributed (iid) noise.

---

[5] In particular, we assume the use of modern, charge-coupled (CCD) cameras, which have replaced photographic film processes as the most dominant imaging form. Nevertheless, most of the observations and models here described equally hold well for other imaging modalities.

[6] i.e., the response at any point in the image depends only on the value of the input at that point and not on the position of the point [8].

2.1.5 Algebraic approach to unconstrained restoration

Following Gonzalez et al. [18], we will focus on an algebraic approach to image restoration (thus considering discrete signals), which has the advantage of allowing the derivation of numerous techniques from the same basic principles. Although direct solution by manipulating vectors and matrices of such a large size is not a trivial task [41], under certain conditions, computational complexity can be reduced to the same level as that required by traditional frequency domain restoration techniques.

2.1.5.1 Unconstrained restoration: Inverse filtering

Due to the random component that ***n*** introduces, the restoration problem translates into obtaining an estimate of the original image given the observed one, according to some predefined criterion of performance.

Specifically, we may seek an $\hat{\boldsymbol{u}}$ such that $\boldsymbol{A(\hat{u})}$ approximates $\boldsymbol{v}$ is some sense. To that end, let $J(\boldsymbol{u})$ be an error metric on the solution space, expressed as the sum of the norm $\rho(\cdot)$ of the residual errors $\boldsymbol{\varepsilon}_i = \mathbf{v}_i - (\mathrm{A}(\hat{\mathbf{u}}))_i$ over the N observation samples. Then, the estimate $\hat{\boldsymbol{u}}$ may be defined so that the objective function $J(\boldsymbol{u})$ is minimized

$$\mathrm{J}(\mathbf{u}) = \sum_{i=1}^{N} \rho\left(\{(\mathbf{v} - \mathrm{A}(\mathbf{u}))\mathbf{Q}\}_i\right) \quad (2.4) \qquad \hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \mathrm{J}(\mathbf{u}) \quad (2.5)$$

If $\rho(\varepsilon_i) := -\log(p_n(\varepsilon_i))$, then $\hat{\boldsymbol{u}}$ is the ***maximum likelihood*** (ML) estimate of $\boldsymbol{u}$, i.e. $\hat{\boldsymbol{u}}$ is the most likely original image giving rise to the observed image $\boldsymbol{v}$.

$$\begin{aligned}
\hat{u} &= \arg\max_{u} p_{\mathbf{v}|\mathbf{u}}(\mathbf{v}\,|\,\mathbf{u}) \\
&= \arg\max_{u} p_{\mathbf{n}}(\mathbf{v} - \mathrm{A}(\mathbf{u})) \\
&= \arg\max_{u} (\log p_{\mathbf{n}}(\mathbf{v} - \mathrm{A}(\mathbf{u}))) \\
&= \arg\max_{u} \sum_{i=1}^{N} \log\{p_n(\{(\mathbf{v} - \mathrm{A}(\mathbf{u}))\mathbf{Q}\}_i)\} \\
&= \arg\min_{u} \sum_{i=1}^{N} \rho(\{(\mathbf{v} - \mathrm{A}(\mathbf{u}))\mathbf{Q}\}_i)
\end{aligned} \quad (2.6)$$

where $\boldsymbol{C_n^{-1} := Q^T Q}$ is the inverse covariance matrix of $\boldsymbol{n}$ (i.e., $\boldsymbol{Q}$ is a so-called decorrelating or whitening matrix)

Finally, remark that, while in the absence of any knowledge about $\boldsymbol{n}$, $\hat{\boldsymbol{u}}$ will not be, in general, an optimal estimate, it can be shown to be asymptotically unbiased and efficient [31].[7]

[7] An important assumption to produce a meaningful unbiased estimate is that the noise and model inaccuracies, n, are zero mean distributed.

2.1.5.1.1 *Least Squares estimate*

Using a quadratic error norm ($\rho(\varepsilon_i) = \varepsilon_i^2$), eq. (2.5) reduces to a weighted least squares (WLS) problem. *J(**u**)* is now the weighted Euclidean distance from ***Au*** to ***v***

$$J(\hat{\mathbf{u}}) = (\mathbf{v} - A(\mathbf{u}))^T W(\mathbf{v} - A(\mathbf{u})) = \| \mathbf{v} - A(\mathbf{u}) \|_W^2 \qquad (2.7)$$

Solving (2.7) with $\boldsymbol{W=C_n^{-1}}$ leads to an optimal estimate (ML) if the noise is Gaussian distributed with covariance matrix $\boldsymbol{C_n}$[8]. However, even if that is not the case, when *A* is a linear operator, least square methods lead to closed form linear solutions that may be computed analytically using the tools of linear algebra.

Linear LS: inverse filtering

If A is a linear operator, then $\mathbf{v} = \mathbf{A} \cdot \mathbf{u} + \mathbf{n}$ and û is also linear

$$\hat{u}_i = \sum_{k \in I} w(i,k) v_k \qquad (2.8) \qquad\qquad \hat{\mathbf{u}} = \mathbf{W}\mathbf{v} \qquad (2.9)$$

Setting the derivative of *J(**u**)* w.r.t. ***u*** equal to zero and solving for ***u***, we get

$$\hat{\mathbf{u}} = (\mathbf{A}^T \mathbf{C}_n^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C}_n^{-1} \mathbf{v} \qquad (2.10) \qquad\qquad \hat{\mathbf{u}} = \mathbf{A}^{-1}\mathbf{v} = \mathbf{u} + \mathbf{A}^{-1}\mathbf{n} \qquad (2.11)$$

Observe that if the noise is iid, then $\boldsymbol{C_n}=\sigma_n \boldsymbol{I}$ and eq. (2.10) reduces to eq. (2.16). For obvious reasons this approach is commonly referred to as ***inverse filtering***.

Finally remark that, while in general the distribution of the noise may remain unknown and, therefore, no claim about the optimality of û can be done, the Gauss-Markov theorem entails that (2.10) is still the best linear unbiased estimator (BLUE) of ***u***.

Ill-posed nature of the restoration problem

While one could think that the restoration problem can now be reduced to simple matrix manipulations, the truth is that direct solution by manipulating vectors and matrices of such a large size is not a trivial task. But even most important is the fact that, although the direct problem (i.e., calculation of the image system response **v** to the true image **u**) is ***well-posed*** in the Hadamard sense[9], the inverse problem is usually ***ill-posed***, when **u** and **v** belong to a Hilbert space. This is because it is ***under-constrained***, i.e. **n** is not known and **A** may be bad-conditioned (causing instability of the solution) or even not invertible. In other words, there is not enough information on **v** to uniquely recover **u**.

---

[8] In this case *J(**u**)* is the ***Mahalanobis distance***, provided that *E[**n**]=0*.

[9] Hadamard (1923) first defined a mathematical problem to be well-posed when its solution *(i)* exists; *(ii)* is unique and *(iii)* depends continously on the inicial data. Notice that for the solution to be robust against noise in practice the problem must be not only well-posed but also well-conditioned. Most of problems of classical physics are well posed.

2.1.6 Regularization model

To force stability and uniqueness of the solution, one commonly restricts the class of admissible solutions to lie in a subspace of the solution space, where it is well defined. This is done by introducing suitable a priori knowledge in the form of generic constraints (such as a smoothness requirement) on the problem. If we lacked such prior knowledge or expectation, nothing better than the degraded image itself could be obtained. Thus, restoration always involves assumptions about the original image.

Regularizing ill-posed problems has been widely investigated in the past (see for example [17], chapter 3.6, for an overview), for which most classical methods proposed fall in two categories: ***(i)*** regularization in functional spaces, and ***(ii)*** control of dimensionality. The framework proposed here belongs to the first category, where the basic idea is to restrict the space of acceptable solutions by choosing the function ***û*** that minimizes an appropriate cost functional (E(u)).
This variational approach to the regularization of the ill-posed problem of finding **u** from **v**, **v**=**Au**+**n**, then requires the definition of a suitable functional space to describe images and their geometrical properties, i.e. the choices of norms $\|\cdot\|$ and of a stabilizing functional $\|P\boldsymbol{u}\|$, dictated by mathematical considerations, and, most importantly, by a physical analysis of the generic constraints on the problem.

2.1.6.1 Standard Regularization

Following [31], we will refer to all techniques that involve the minimization of a quadratic functional as ***standard regularization*** theory (Tikhonov), where A and P are linear operators and the norms are quadratic. Then, the three main methods that yield an estimate û are: ***(i)*** among u that satisfy $\|P\boldsymbol{u}\| \le C$, find u that minimizes $\|\boldsymbol{v}\text{-}A\boldsymbol{u}\|^2$; ***(ii)*** among u that satisfy $\|\boldsymbol{v}\text{-}A\boldsymbol{u}\| \le \varepsilon$, find u that minimizes $\|P\boldsymbol{u}\|^2$; and ***(iii)*** find u that minimizes a weighted average of the previous terms

$$\mathrm{E}(\mathbf{u}) = \underbrace{\| \mathbf{v} - \mathbf{A}\mathbf{u} \|^2}_{similarity} + \lambda \underbrace{\| \mathrm{P}\,\mathbf{u} \|^2}_{smoothness} \qquad (2.12) \qquad \hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \{\mathrm{E}(\mathbf{u})\} \qquad (2.13)$$

where λ ($\lambda=\varepsilon/C$), a so-called regularization parameter, controls the compromise between the degree of regularization of a solution (second term, often called ***regularizer*** or ***smoothness*** -or ***roughness- term***[10]) and its closeness to the observed image (first term, often called ***similarity*** or ***fitting term***). The problem of selecting the optimal value of λ for image restoration is discussed in some detail in Chapter 4. These equations form the bases for all restoration procedures discussed in the following sections.

---

[10] One of the most elementary global regularizers is smoothness. For many problems in image processing it makes sense to demand that a quantity to be modeled changes only slowly in space and time [19]. The smoothness assumption (limited band signal) gives rise to high-pass penalty operators; so, in this case the unwanted feature is the energy in the high-frequency region.

Marroquin et al. showed in [31] that several problems in early vision such as surface reconstruction, optical flow and stereo can be solved using standard regularization techniques. In addition, parallel architectures and analog networks have been proposed as natural implementations.

Solving for û

In standard regularization, the estimate û is linear, i.e. the restored value at each point can be written as a linear combination (weighted average) of all the values of the image:

$$\hat{\mathbf{u}} = \mathbf{W}\mathbf{v} \qquad (2.14) \qquad\qquad \hat{\mathbf{u}}_i = \sum_{j \in I} w(i,j)\mathbf{v}_j \qquad (2.15)$$

Using linear algebra to solve (2.13) yields the explicit solution

$$\mathbf{W} = (\mathbf{A}^T\mathbf{A} + \lambda \cdot \mathbf{P}^T\mathbf{P})^{-1}\mathbf{A}^T \qquad (2.16)$$

which reduces to the non-regularized least squares solution (commonly known as *inverse filtering*) when $P=0$, provided that $(\mathbf{A}^T\mathbf{A})^{-1}$ exists.

2.1.6.2 Relation to Diagonal operators: Wiener Filter and SVD

If $P=C_u^{-1}C_n$, then the classical Wiener solution is obtained, i.e. $\hat{u}=E(u \mid v)$, which minimizes $E(|u-\hat{u}|^2)$. In case $P=I$ then the *parametric* Wiener filter.

Given the Singular Value Decomposition (SVD) $A = USV^T$ with singular values s_i, and $P=I$, the least squares solution in (2.16) can be expressed as $\hat{u} = VDU^T v$, with singular values d_i given by

$$d_i = \underbrace{\frac{s_i^2}{s_i^2 + \lambda}}_{w_\lambda(s_i^2)} \frac{1}{s_i} \qquad (2.17)$$

where the filter function $w_\alpha(s_i^2)$ shows that Thikonov regularization actually filters out singular components that are small (relative to λ) while retaining components that are large.

This leads to the second family of regularization methods: the well-known truncated singular value decomposition (TSVD) (see [17], chapter 3.6) where regularization is performed through control of the dimensionality of the solution space. Here the summation in (2.14) is taken on a limited number of singular values s_i that significantly differ from zero, in order to avoid excessive noise contamination. This is equivalent to projecting the solution onto a "significant" subspace spanned by the remaining singular vectors. This method yields solutions that are numerically well conditioned, while providing a “sharp cut-off” behavior instead of “smooth roll-off behavior” of Tikhonov filter.

2.2 Vision Models

The visual system processes information at many levels of sophistication. At the retina, there is low-level vision, including *light adaptation* and the *center-surround* receptive fields of ganglion cells. At the other extreme is high-level vision, which includes cognitive processes that incorporate knowledge about objects, materials, and scenes. In between there is mid-level vision. Mid-level vision is simply an ill-defined region between low and high [30][33].

The low-level approach to vision is mechanistic, in that it seeks to formulate mathematical models that incorporate known features of early stages of the neural processing chain. It is associated with Ewald Hering, who considered adaptation and local interactions, at a physiological level, as the crucial mechanisms. This approach has long enjoyed popularity because it offers an attractive connection to ***psychophysics***, which abstracts further and considers the problem as one of characterizing the properties of a black-box system, where the effort is on testing principles that allow prediction of the relationship between physical measurements of a stimulus and the perception of that stimulus [33]. Colour appearance is a good psychological domain for working out the link between psychological and neural results because several quantifiable phenomena such as ***trichromacity***, ***gain control*** or ***light adaptation, opponent-colour encoding***, and ***low spatiotemporal resolution for colour***, are well-characterized. The work in [33] and [34] surveys these and other well-established results from physiology and psychophysics about early vision that are important for computer graphics but often overlooked by. Because of its relevance to support the ideas in this thesis, we reproduce an excerpt in section 2.5.1 *Visual Psychophysics*.

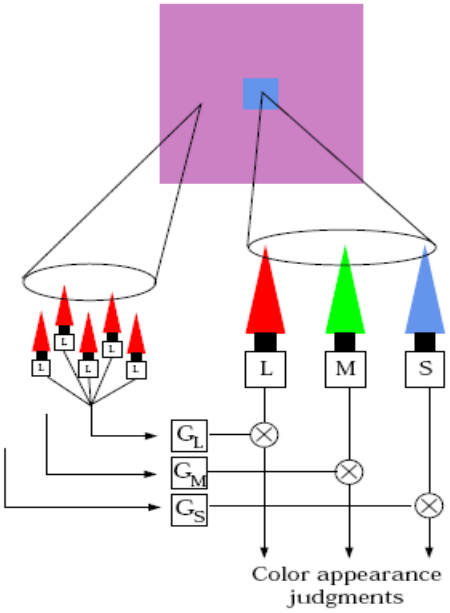


Figure 2.3. Mechanisms of colour appearance: ***(a)*** trichromacy (there are three perceptual dimensions); ***(b)***, adjustment to changes in the ambient illumination, so that appearance depends more on the relative rather than absolute cone absorptions; ***(c)***, organization into one achromatic and two opponent-colours representations.

The effectiveness of such bottom-up approaches depends on how much the HVS is understood and how accurately the simulation can be implemented. While much of our knowledge about the visual system as come through these two approaches, the fact that their results are rather limited (they respectively apply to very low-level neural processing and to rather simplified visual conditions) on the one hand, and descriptive but not explanatory on the other, presents a major drawback for their application in real-world problems.

2.2.1 Computational Vision

The high-level approach, historically associated with Helmholtz[11], is computational in the sense defined by Marr [30]. The idea is here to develop models by thinking about what useful function vision serves and how a system designed to accomplish this function would perform. E.g., in the case of colour constancy, one asks how could a visual system process image data to recover descriptions of object surface reflectance that are stable across changes in the scene.

Vision is an information-processing task with well-defined input (the sensory data) and output (a concise description of the scene depicted in the image, the exact nature of which depends upon the goals and expectations of the observer [26]). In order to better understand how the later is computed from the former, it is customary subdivide the processing of images by the visuo-cognitive system in three distinct processes [26]: *(1)* perception, that is, the construction of an internal representation of the image using primarily low-level knowledge of the visual world; *(2)* interpretation, that is, the confrontation ("matching") of this internal representation with memory representations; and *(3)* task-directed semantic processing of the interpreted scene in order to formulate a response (see Figure 2.5)

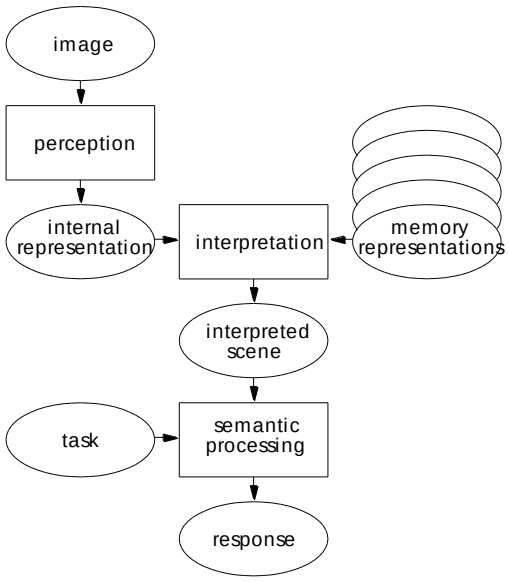


Figure 2.4. A diagrammatical depiction of visuo-cognitive processing of images. Reproduced from [43].

Goals and knowledge are very important high-level capabilities that can guide visual activities, and a (visual system) should be able to take advantage of them. They are, however, only a part of the vision story. Vision requires many low-level capabilities we often take for granted; for example, our ability to extract intrinsic images of "lightness", "colour", and "range". Such capabilities are elusive, unconscious, and not well connected to other systems that allow direct introspection. Skipping the low-level processing we take for granted turns normally effortless perception into a very difficult puzzle. Computer vision is vitally concerned with both low-level or "early processing" issues and with the high-level and "cognitive" use of knowledge [25]. In this thesis, we postpone to the last chapter high-level internal model information even though it is important and can affect early processing.

[11] In contemporary vision research, the dominant theoretical approach can be traced to Helmholtz's constructivism (what he terms the empirical theory), which is the view that we construct internal representations of the objects in the visible scene that are most likely to have given rise to the pattern of light concurrently impinging on our retinas.

2.2.2 Early Vision: recovering intrinsic components

The first task that a visual system has to solve consists in recovering, from the set of two-dimensional images that constitute the sensory input, relevant physical properties of the visible three-dimensional world, such as colour, shape, range, orientation, reflectance, and incident illumination. Barrow and Tenenbaum [26] proposed using ***intrinsic images*** to represent these characteristics. For a given intrinsic (vertical) characteristic of the scene, the pixel of the corresponding intrinsic image would represent the value of that characteristic at that point.[12]

Support for this idea comes from three sources: ***1)*** the obvious utility of intrinsic characteristics as a stepping-stone to higher level scene analysis and perceptual operations, ranging from segmentation to object recognition, that have so far proved difficult to implement reliably; ***2)*** the apparent ability of humans to determine these characteristics, regardless of viewing conditions or familiarity with the scene (e.g., shadows are usually easily distinguished from changes in reflectance); and ***3)*** a theoretical argument that such a description is obtainable, by a non-cognitive and non-purposive process, at least, for simple scene domains.

The recovery of intrinsic scene characteristics is a plausible role for early stages of visual processing (usually called ***early*** or ***low-level vision***), where it is assumed to be performed (in natural systems) by a set of generic processes that correspond to conceptually independent computational modules, each one specialized in the reconstruction of a particular property, that can be studied, at least to a very first approximation, in isolation. Specific early vision modules have been proposed for the computation of: brightness edges; surface colour, lightness and albedo; shape and depth from contours, texture, shading, stereo and motion; velocity and optical flow; etc.

The central problem in recovering intrinsic scene characteristics is that the information is confounded in the original light-intensity image: photometrically, the light intensity at each point in an image can result from an infinitude of combinations of illumination, reflectance, and orientation at the corresponding scene point; geometrically, the distance at each point in the image is lost in the projection from the three-dimensional world, resulting in the kinds of ambiguities depicted in Figure 2.5. Recovery is thus an under-constrained problem that requires additional constraints for solution. Therefore, recovery depends on exploiting constraints, derived from assumptions about the nature of the scene and the physics of the imaging process.

---

[12] An image of a scene can be modelled as the composition of a scene's intrinsic component images. At the image formation level, the image can be represented as the sum of a noise image and the true image of the scene. The scene image itself can also be represented as the composition of images that describe the characteristics of the surfaces in the scene (e.g., a shading image, which describes the illumination and shape of the surface, and an albedo image, which describes the reflectance of the surface. See Chapter 5).

2.2.3 Bayesian formulation of visual perception

In this section, we approach perception as a visual *inference problem* and, more specifically, as *statistical inference*, which has long provided a common framework for modelling artificial and biological vision [34]. This means that, instead of thinking of neural representations as transformations of stimulus energies, we will regard them as approximate estimates of the probable truths of hypotheses about the current environment.

An attractive general principle is that vision resolves ambiguity by taking advantage of the statistical structure of natural scenes: e.g., given several physical interpretations that are consistent with the sense data, the visual system may choose the one that is most likely a priori [15][32]. This principle can be formalized using Bayesian decision theory, and there is great interest in linking perceptual performance explicitly to Bayesian models [28][29].[13]

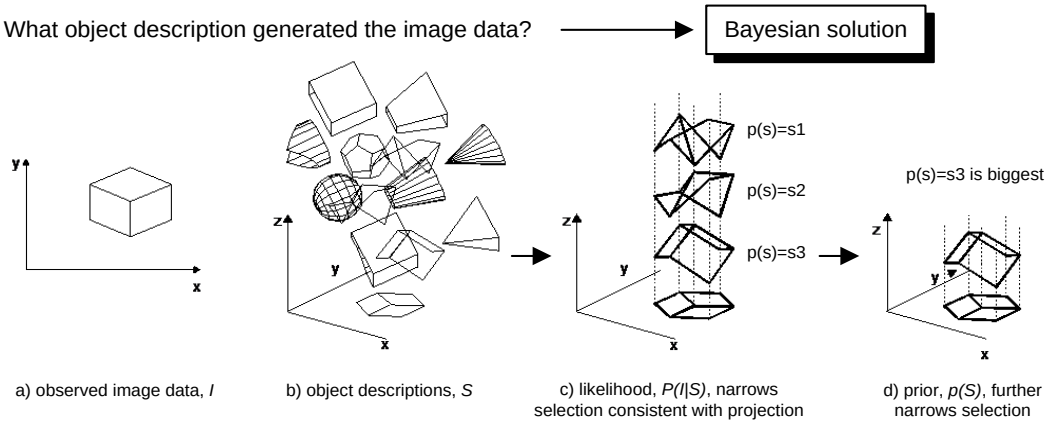


Figure 2.5. Illustration of Bayesian theory applied to object perception: *(a)* What 3D object caused the image of a cube? The likelihood *p(I|S) = p(image data|object descriptions)* constrains the possible set of objects to those consistent with the image data, but even this is an infinite set. *(b)* The prior knowledge probability *p(S) = p(object descriptions)* constrains the consistent set of 3D objects to those that are more probable in the world. *(c,d)* The probability over all instances is determined by the product of the likelihood and prior knowledge: *p(S,I) = p(I|S)p(S) = p(object descriptions, image data)*. Adapted from [28].

2.2.3.1 Information for Inference

One can identify three types of constraints that make reliable visual inference possible: the visual task, prior knowledge of scene structure independent of the image, and the relationship between image structure and task requirements. Bayesian decision theory provides a precise language to model these constraints [29]. We postpone discussion of the visual task, and suppose that the image measurements (e.g., the image intensity values or features extracted from them), **I**, and the required scene parameters (e.g., vector with variables representing surface shape, material reflectivity, illumination direction, and viewpoint), **S**, useful for the task have already been specified.

[13] Two important different frameworks for understanding the human visual system in a fundamentally statistical manner are Bayesian decision theory and empirical ranking theory [32].

The knowledge for visual inference is characterized by the posterior probability distribution, $p(\mathbf{S}\,|\,\mathbf{I})$ which models the probability of a scene description **S**, given the observed image data, **I**. By Bayes' rule, the posterior is:

$$p(\mathbf{S}\,|\,\mathbf{I}) = p(\mathbf{I}\,|\,\mathbf{S})\,p(\mathbf{S}) / P(\mathbf{I}) = C \cdot p(\mathbf{I}\,|\,\mathbf{S})\,p(\mathbf{S}) \qquad \textit{(eq.18)}$$

where C is a normalizing constant, $p(\mathbf{S})$ is the prior probability distribution modelling the scene (i.e., the information we have about **S** before observing **I**) and $p(\mathbf{I}\,|\,\mathbf{S})$ is the likelihood distribution modelling image formation.[14] Roughly this equation reads: *the reliability of the information provided about some scene property given an image p(S/I) is equal to the likelihood of obtaining that image given the scene p(I/S) scaled by a measure of how often that scene property occurs p(S). The denominator p(I) is a normalizing constant*.

2.2.3.2 Vision by an Agent: Decision Theory

While posterior distribution completely defines the visual information available, most often one still has to extract estimates of scene parameters and make decisions according to some criteria. Generally, an optimal perceptual decision $\mathbf{S}' = \alpha(\mathbf{I})$ is a function of the task as well as the posterior. Each task can be associated an appropriate ***loss function***, $L(\mathbf{S}',\mathbf{S})$, which specifies the cost (*penalty*) of guessing (*deciding*) $\mathbf{S}'$ when the actual scene variables are **S**. Then the cost or ***Bayesian risk***, $R(\mathbf{S}',\mathbf{I})$, associated with each possible interpretation of the stimulus, is defined as the expected loss (or the negative of the expected utility), taken with respect to the posterior distribution, $p(\mathbf{S}\,|\,\mathbf{I})$:

$$R(\mathbf{S}',\mathbf{I}) = \sum_{\mathbf{S}} L(\mathbf{S}',\mathbf{S})\,p(\mathbf{S}\,|\,\mathbf{I})$$

The Bayes ideal observer then picks the interpretation with minimum risk (i.e., maximum expected utility) [29].[15]

Lost functions - Risk - Cost (task specification), estimation

Some tasks require an observer to maximize the proportion of correct decisions, which translates into a *minus-delta* loss function, i.e. $L(\mathbf{S}',\mathbf{S}) = -\delta(\mathbf{S}'-\mathbf{S})$. In this case, the risk becomes $R(\mathbf{S}',\mathbf{I}) = -p(\mathbf{S}'\,|\,\mathbf{I})$, and then the best strategy is to choose the scene description $\mathbf{S}'$ for which the posterior is biggest. This is known as ***MAP*** (Maximum A Posteriori) estimator. Other tasks require that $L(\mathbf{S}',\mathbf{S}) = (\mathbf{S}'-\mathbf{S})^2$, resulting in the (Bayesian) ***MMSE*** (Minimum Mean Square Error) or ***Bayes LS*** estimator. However, it has been noticed that neither the minus-delta nor the squared-error cost functions are appropriate in perception problems, where an estimate of the scene parameters that is *approximately* correct will often do, and once the estimation error is sufficiently large, the loss may saturate. Instead, the ***local mass loss function*** is to be preferred.

---

[14] The generative model, $\mathbf{S}\rightarrow\mathbf{I}$, may be thought of as the *rendering equation*, describing the process of generating an image from a description of 3D objects, expressed as a probability distribution.

[15] Theoretical observers that use Bayesian inference to make optimal interpretations are called *ideal observers*. An ideal observer does not necessarily get the right answer for each input stimulus, but it does make the best guesses so it gets the best performance averaged over all the stimuli. In this sense, an ideal observer may "see" illusions.

Discounting and Task Dependence: explicit and generic variables

Often, we can simplify the task requirements by splitting **S** into components ($\mathbf{S}_1$; $\mathbf{S}_2$) that specify which scene properties are important to estimate ($\mathbf{S}_1$, e.g., surface shape) and which confound the measurements and are not worth estimating ($\mathbf{S}_2$, e.g., viewpoint, illumination). These are commonly referred to in the literature as ***explicit*, or *intrinsic*** -to the scene or object-, and ***generic*, *extrinsic*** or ***confounding*** variables, respectively. Confounding variables are analogous to noise in classical signal detection theory, but they are more complicated to model and they affect image formation in a highly nonlinear manner. For example, a standard noise model has $\mathbf{I} = \mathbf{S} + \mathbf{n}$, where **n** is Gaussian noise. Realistic vision noise is better captured by $p(\mathbf{I}|\mathbf{S})$. Here, the problem is making a good guess independent of (or invariant to) the true value of the confounding variable. The task itself can serve to reduce ambiguity by discounting the confounding variable [28]. From the Bayesian perspective, we discount the confounding variables by integrating them out (or summing over them). This marginalization of the posterior with respect to the confounding variable, which means that costs are constant over all guesses of it, is equivalent to treating the variable as having such low utility that it is not worth estimating.

$$p(\mathbf{S}_1, \mathbf{I}) = \sum_{\mathbf{S}_2} p(\mathbf{S}_1, \mathbf{S}_2, \mathbf{I})$$

2.2.3.3 Integration of Image Measures and Cues: feature reliability

Vision integrates information from a variety of sources. For example, one can identify more than a dozen cues that the human visual system utilizes for depth perception. This can be modelled from the Bayesian perspective by considering the *reliability* of each cue. When the variables are Gaussian and conditionally independent given the shared explanation, and we have estimates for each cue alone (i.e., S_i' is the best estimate of S_i from $p(S|I_i)$), then optimal integration (i.e., the most probable value) of the two estimates considers the uncertainty owing to measurement noise and is given by the weighted average[16],

$$S' = S_1' \frac{r_1}{r_1 + r_2} + S_2' \frac{r_2}{r_1 + r_2}$$

where r_i, the reliability, is the reciprocal of the variance[17]. This model has been used to study whether the human visual system combines cues optimally. Prior probabilities and likelihoods can also combine like weighted cues. Under this view, Bayes formula implies that perception is a trade-off between image feature reliability, as embodied by the likelihood $p(\mathbf{I}|\mathbf{S})$, and the prior probability $p(\mathbf{S})$. Some perceptions may be more *prior driven*, and others more *data driven*. The less reliable the image features (e.g., the more ambiguous), the more the perception is influenced by the prior. This trade-off has been illustrated for a variety of visual phenomena.

---

[16] A more complicated model uses robust statistics to determine whether one measurement is an outlier, and therefore should not be integrated with the other measurement.

[17] It is a general property that whenever two independent sources contribute information via Gaussian distributions about an unknown variable, the precisions add.

2.2.3.4 Perceptual Explaining Away

The term 'explaining away' originates in the context of reasoning where a change in the probability of one competing hypothesis affects the probability of another. From a Bayesian perspective, the competition results from the two (otherwise independent) hypotheses becoming conditionally dependent given the image (then lowering the probability of one of the explanations increases the probability of the other). It has been argued that this accounts for a range of visual phenomena, including the estimation of material properties. A striking example of perceptual '*explaining away*' is shown in Figure 2.6.

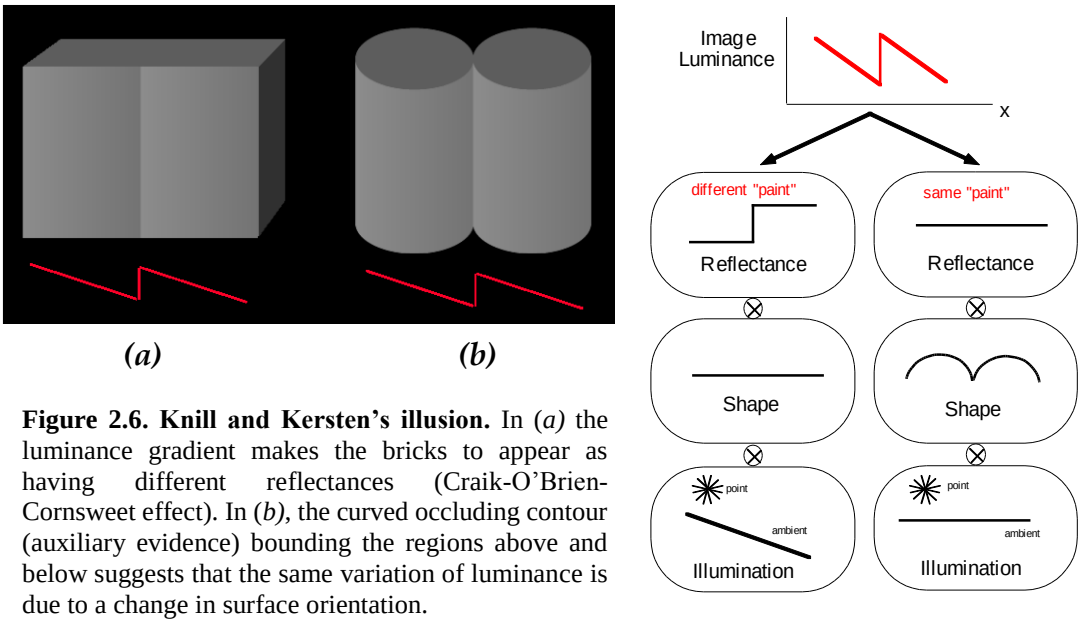


Figure 2.6. Knill and Kersten's illusion. In (*a*) the luminance gradient makes the bricks to appear as having different reflectances (Craik-O'Brien-Cornsweet effect). In (*b*), the curved occluding contour (auxiliary evidence) bounding the regions above and below suggests that the same variation of luminance is due to a change in surface orientation.

2.2.3.5 Difficulties

Without direct input, how does image independent knowledge of the world get built into the visual system? One possible answer is that the priors are coded in the genes, as probability estimates derived from the frequencies of survival and death involved in natural selection. Generalization capabilities (especially in children) allow their further improvement within the experience of a particular individual. Psychophysical experiments can test theories regarding knowledge specified by $p(\mathbf{S})$.[18]

Recent studies show considerable statistical regularities in natural images and scene properties that help tame the problems of complexity and ambiguity in ways that can be exploited by biological and artificial visual systems. For certain problems it is also possible to learn the posterior distribution $p(\mathbf{S}|\mathbf{I})$ directly, which relates to directly learning a classifier $\alpha(\mathbf{I})$. Some authors have shown

[18] One of the best-known examples of a prior is the assumption that light is coming from above. This assumption is particularly useful to disambiguate convex from concave shapes from shading information. The light from above prior is natural when one considers that the Sun and most artificial light sources are located above our heads [28].

however, that under certain conditions it is possible to formulate the Bayesian least squares (BLS) approach without explicit prior (see Chapter 4).

The theoretical difficulties of the Bayesian approach reduce to two issues. First, can we learn the probability distributions p(**I** | **S**) and p(**S**) from real data? Second, can we find algorithms that can compute the best estimators?

2.2.3.6 Conclusions: The importance of vision as Bayesian inference

The Bayesian approach yields a uniform framework for studying object perception. It distinguishes itself from other statistical formulations by taking into account the contributions of both the image formation process and the statistical structure of the world to the specification of the available information. In particular, the approach is notable for its reliance on explicit models. While this forms the basis for most attacks on the approach, it must be emphasized that modelling this aspect of visual information is a fundamental necessity, and is always implicitly done, if not explicitly.

As pointed out in [28], the benefits of this Bayesian framework are that: 1) it explicitly models uncertainty. This is important in accounting for how the visual system combines large amounts of objectively ambiguous information to yield perceptions that are rarely ambiguous; 2) it provides a principled way to choose an optimal estimate that uses all of the information contained in the data; 3) it allows the development of quantitative theories at the information processing level, avoiding premature commitment to specific neural mechanisms; and 4) it ties naturally to theories of perception and cognition involving top-down feedback or *analysis by synthesis*. This means that high-level hypothesis regarding objects properties could be used at low-level stages to resolve ambiguities in the incoming retinal image measurements. Notice, however, that Bayesian models can be applied to every stage in the vision chain, such as object perception or surface perception. We focus on the latter, since it belongs to early-vision.

2.2.3.7 Connection to regularization.

Bayesian methods may be used to understand non-Bayesian algorithms. The key idea is to find a prior and a loss function for which the non-Bayesian algorithm mimics the Bayesian solution. For example, while inverse problems have traditionally been solved by energy minimization or regularization methods, a better understanding can be obtained by replacing these methods with Bayesian ones to get explicit modelling of uncertainty. This has lead to the so-called *stochastic* or *probabilistic* approach to regularization, which became popular in image restoration with the seminal work of Geman and Geman [91]. This is the spirit followed in Chapter 4 to introduce probabilistic approaches to feature-preserving image smoothing based on a priori assumptions about both the image structure and the distortion process.

2.3 Overview of Image Modeling

Since the retinal image is often ambiguous, the visual system's success in interpreting images must be because it makes god assumptions about likely properties of objects in the world. Among all the possible combination of colourant arrangement that one could imagine, ***natural images*** is the term commonly used to refer the very small set which results from optical registration from the real visual world similar to the one performed by the eye or a camera. This differentiates them from text, computer graphics, cartoons, paintings, drawings, random patterns, images derived from invisible radiation, etc. Consequently, the information contained in natural images manifests itself, virtually always, in some patterns evident in the image data. We refer to these patterns as the regularity in the data [40]. Understanding and describing this regularity in a way that is both general and powerful is one of the key problems in vision science as well as in image processing. We refer to the process of describing regularity in images as image modeling. Image models play a fundamental role as *a priori* knowledge in source coding, estimation and decision problems. Researchers have taken different kinds of image modeling approaches including those based on (a) *geometry*, (b) *statistics*, and (c) *wavelets* [17][18][19][20][21]. We briefly describe the characteristic features of each of these models.

2.3.1 Variational image modelling

Geometric image modeling relies on the interpretation of an image as a function defined on a grid domain, describing and analyzing the local spatial relationships (or geometry) between the function values via tools relying on calculus. This invariably connects to the fields of differential geometry and differential equations, treating images as functions that can be considered as points in high-dimensional Sobolev spaces[19]. Modeling image functions in such spaces, however, does not accommodate for the existence of discontinuities, or edges, in images. Edges are formed at the silhouettes of objects and are vital features in image analysis and processing. To accommodate edges in images, two popular models have been proposed: ***a)*** the ***object-edge model*** (invented by Mumford and Shah [37]), which assumes that the grid image domains can be partitioned into mutually-exclusive and collectively-exhaustive sets such that the resulting functions on each partition belong to Sobolev spaces; and ***b)*** the ***bounded-variation image model*** (proposed by Rudin, Osher, and Fatemi [39]), where images are assumed to possess bounded variation. Both these image models, however, impose strong constraints on the data and do not apply well to textured images. To explicitly deal with textured images, researchers have proposed more sophisticated image models, known as ***cartoon-texture models***, which decompose an image into the sum of a piecewise-constant part and an oscillatory texture part.

---

[19] A *Sobolev space* is a normed space of functions such that all the derivatives up to some order k, for some $k \geq 1$, have finite L_p norms, given $p \geq 1$.

2.3.2 Statistical modeling in the image domain

Statistical models, on the other hand, aim to capture the variability and dependencies in the data via joint or conditional PDFs. Specifically, they treat image data as realizations of random fields. Such models are good at capturing the regularities in natural images that are rich in texture-like features [23].

Modeling the statistics of natural images is a challenging task, partly because of the high dimensionality of the signal. However, the observation that natural images present a strong correlation between the luminance of pixels in the spatial domain, has led to two common assumptions adopted in order to reduce the high dimensionality of the image signal, as described in their works by Portilla and Simoncelly (see for example [38]): 1) ***locality*** i.e., the probability density of a pixel is independent of the pixels beyond its neighbourhood (Markov assumption)[20]; and 2) ***homogeneity, i.e.,*** the distribution of pixel values in a neighborhood is the same for all such neighborhoods, regardless of absolute spatial position within the image. This is a translation-invariance assumption.

Scale-invariance assumption (resizing the image does not change the probability structure) and translation-invariance assumptions, on which autocovariance characterization relies, along with gaussianity, lead to consider images as samples of a Gaussian random field (GRF) with variance falling as f^{-g} in the frequency domain, as shown by a number of empirical studies. Under the Gaussian assumption, a simple second-order approach fully describes the signal. This simple characterization has been successfully exploited in applications where removing the statistical dependence of the samples is required [36]. E.g., principal component analysis (PCA) on a set of natural images gives rise to Fourier-like eigenvectors, i.e., oscillating functions extending all over the spatial domain. Moreover, the energy (the square of the eigenvalues) is concentrated in the low-frequency PCA components [41]. Specifically, the f^{-g} power spectra justifies the use of high-pass regularization operators, such as first and second derivatives of the signal, in image restoration.

While, this suggests that statistical models can be further improved by selection of appropriate image representation in order to reduce the spatial correlation (e.g., Fourier, PCA, DCT, etc. transformations), natural images are not that simple though. In order to consider higher order interactions, independent component analysis (ICA) techniques have been developed as an alternative to PCA. When applied to natural images, wavelet-like representations emerge, i.e., spatially localized oscillating functions.

[20] This poses images as realizations of a Markov random field (MRF), in which the conditional probabilities for image neighborhood configurations, namely cliques, encode a set of probabilistic assumptions (priors) about the geometric properties of the signal. Theoretical and applied research over the last few decades has firmly established MRFs as powerful tools for statistical image modeling and processing.

2.3.3 Statistical modelling in the transform domain: Wavelet

From yet another perspective, images are formed as a superposition of local responses from some kind of sensor elements. Moreover, they exhibit such phenomena at multiple scales [40]. These local dependencies at multiple scales are well captured, mathematically as well as empirically, by the wavelet-based models [17][18][19][20][21]. Some limitations of these methods stem from the choice of the particular wavelet decomposition basis as well as the parametric models typically imposed on the wavelet coefficients.

Over the past decade, it has become standard to initiate computer-vision and image processing tasks by decomposing the image with a set of multiscale bandpass oriented filters that enable scale-space-orientation analysis. This kind of representation, loosely referred to as wavelet decomposition, is selectivity in both, spatial and frequency domain, and effective at decoupling the high order statistical features of natural images [38]. In addition, it shares some basic properties of neural responses in the primary visual cortex of mammals which are presumably adapted to efficiently represent the visually relevant features of images. This is known as the Barlow hypothesis, in which he argued that biological vision systems have evolved for an optimal processing of natural images [33]. A number of results support this hypothesis. First, it has been shown that the early processing mechanisms in the visual cortex perform a linear wavelet-like transform using a set of filters similar to those obtained by applying independent component analysis (ICA) to a set of natural images [26][27]. Second, biological vision systems exhibit nonlinear interactions between the responses of the linear wavelet-like stage. In these nonlinearities, each coefficient is normalized by a combination of neighbouring coefficients. Moreover, greater similarity and flexibility can be achieved by redundant pyramidal representations, which also reduce the artifacts, at the price of increasing sample inter-correlation (this is due to the fact that more samples are considered without addition of information).

Further studies have reported that wavelet coefficients of wavelet image representation exhibit a typical non-Gaussian behavior and high order statistical dependencies not eliminated through decorrelation[21], concluding that the previously mentioned second order characterization would be inadequate. Some authors [38] have proposed the use of Gaussian scale mixtures (***GSM***s) in the wavelet domain for modeling the statistical behavior of natural images. The GSM framework [36] can model the marginal statistics of the wavelet coefficients, the nonlinear dependencies between them, as well as the space-varying localized statistics, proving promising results in several image processing fields such as image restoration or enhancement.

[21] Notice that decorrelation ⇔ independency just holds for the Gaussian case.

2.3.3.1 Note 1: Overview of Gaussian scale mixtures

Briefly, a Gaussian scale mixture is obtained by adding up a continuum (as opposed to the classical Gaussian mixtures, ***GM***s) of zero-mean Gaussian densities, each one with a variance proportional to z, and with a weight given by $p_z(z)$. The resulting distribution is always leptokurotic (kurtosis ≥ 3). Formally, a GSM is described as the product of a hidden positive scalar random variable ($\sqrt{z}$) times a zero mean Gaussian vector (u), $x = \sqrt{z}u$ 0, with density

$$p_x(x) = \int_0^\infty p_{x|z}(x|z)p_z(z)dz = \int_0^\infty \frac{\exp\left(-x^T(zC_u)^{-1}x/2\right)}{(2\pi)^{N/2}|zC_u|^{1/2}} p_z(z)dz \qquad (2.19)$$

where C_u is the covariance matrix of u, and $p_z(z)$ is the multiplier density

2.3.3.2 Note 2: Parametric vs. Nonparametric Statistical Modeling

Broadly speaking, a statistical model is a set of probability density functions (PDFs) on the sample space associated with the data. Parametric statistical modeling parameterizes this set using a few control variables. An inherent difficulty with this approach is to find suitable parameter values such that the model is well-suited for the data. Nonparametric statistical modeling fundamentally differs from this approach by not imposing strong parametric models on the data. It provides the power to model and learn arbitrary (smooth) PDFs via data-driven strategies. As we will show in Chapter 4, non-parametric schemes -that adapt the model to best capture the characteristics of the data and then process the data based on that model- can form powerful tools in formulating unsupervised adaptive image-processing methods.

2.4 Summary

In the field of image processing, we have focused on the classical restoration problem, where the power of the algebraic approach is evident in the simplicity by which methods such as Wiener and constrained least squares filters can be obtained. Most of the restoration techniques derived in preceding sections are based on a least squares criterion of optimality. The use of the word optimal in this context refers strictly to a mathematical concept, not to optimal response of the human visual system. While recent work in information throughput based restoration is beginning to suggest new mathematical approaches to spatial processing, the research could be better understood, if not unified, upon the basis of an understanding of human vision. In fact, the symbiotic relationship developed between the study of digital image processing and of the human visual system holds much promise for the advent of both areas.

Human visual models provide us with a unifying basis for our understanding of the visual process itself as well as for the application of this knowledge to the processing of images (e.g., for image compression, quality assessment and enhancement). Aside from the direct results such applications bring, human visual modeling aids our understanding of image processing problems by providing us with valuable analogies. We focus on low-level (a.k.a *early* vision), where the objective ambiguity of images arises if several different scene features could have produced the same image description. In this case, the visual system is forced to guess, but it can make intelligent guesses by biasing its guesses toward typical objects or interpretations.

From this point of view, an "image processor", like so much the visual system, must exploit the ecology of images, i.e., it must "know" the likelihood of various things in the world, and the likelihood that a given image-property could be caused by one or another world-property. This world-knowledge may be hard-wired (i.e., coded) or learned, and may manifest itself at various levels of processing. Recent work in Bayesian theories of visual perception has shown how complexity may be managed and ambiguity resolved through the task-dependent, probabilistic integration of prior knowledge about likely physical configurations of the world with the information contained in the image.

Recent studies show considerable statistical regularities in natural images and scene properties that help tame the problems of complexity and ambiguity in ways that can be exploited by biological and artificial visual systems. From this perspective, we expect that an image-processing algorithm based on human vision will provide a good solution to the problem. Indeed, as our knowledge of the human visual process grows, more of its complexity and its adaptive nature will surely be modeled. This will lead to the development of smarter image processors which will be able to consistently extract and process desired image information under a wide range of image-forming conditions.

2.5 Appendix

2.5.1 Visual Psychophysics

Psychophysics is the scientific study of the stimulus-sensation relationships (e.g., relationships between physical amounts of light -stimulus- and perceived brightness -perception-).[22] The work in [33] and [34] surveys well-established results from physiology and psychophysics about early vision that are important for computer graphics but often overlooked by. Because of its relevance to support the ideas in this thesis, we reproduce here an excerpt.

2.5.1.1 Psychophysical methods: Threshold and scaling. A historical perspective on Weber, Fechner, and Stevens

The field of psychophysics had its origins in the nineteenth century, when experimentation by Weber and others sought to relate discriminable differences in sensation to continuos physical properties such as weight. These experiments tended to show that, for a given starting weight, I, the change in weight necessary to elicit a perceptual difference, ΔI, followed a constant ratio $\Delta I/I$. This stimulus change is often referred to as a *just noticeable difference*, or a JND.[23] This simple relationship was found to hold approximately true for many different stimuli and has since become known as *Weber's law*.

In 1860, Fechner proposed a method for extending Weber's law to create a quantitative scale of sensation, allowing perceptions to be mapped to numerical values. By integrating over that equation, for all stimuli I, it is possible to calculate a metric that equates equal ratios on the physical scale with equal increments on the perceptual scale. This solution ends up being a simple logarithmic relationship, $S = k \log (I)$, where S is the perceived sensation, k is some constant, and I is the measured physical intensity. This solution became known as *Fechner's law*. The logarithm expressed by Fechner's law represents a compressive nonlinear relationship between the input stimulus intensity and the corresponding perceptual sensation. Fechner's law relies on several fundamental assumptions. First, it assumes that Weber's law is indeed valid for all stimulus intensity (in the limit, $\Delta I/I$ is a constant). His other assumption is that JNDs are indeed a valid unit of sensation and that JNDs can be integrated to form a magnitude scale. While the general compressive trends described by Fechner's law are often valid for many perceptions, they often do not follow the exact logarithmic shape. Perhaps, because the two main assumptions often break down in real-world situations, so Fechner's law is not always accurate.

---

[22] Psychophysics can be used to generate quantitative measurements of color sensation and perception, though those are often thought of as being very subjective. These measurements of perception, when produced from a carefully designed experiment, are just as objective as any other physical measurement (such as temperature). The difference between physical and psychophysical measurements tends to lie in the uncertainty of those measurements [21].

[23] The perceptual uncertainty, reflected in the lack of a deterministic outcome in comparisons of very similar stimuli, can be related to the extremely useful concept of a just noticeable difference (JND).

Nearly 100 years later, S. Stevens performed a series of experiments testing the limits of Fechner's law. It was found that most of the relationships formed straight lines when plotted on a log-sensation/log-intensity plot, rather than the logarithmic relation predicted by Fechner's law. From these plots, Stevens suggested that the relationships between physical stimuli and their corresponding perceptual scales could be defined as power functions, where the exponents vary for different perceptions. The general form of this is $S=kI^{\gamma}$. An exponent greater than 1 results in an expansive relationship; as the physical stimulus increases, the perception increases at a greater rate. This is often the case when the stimulus might result in danger, such as the perception of pain. An exponent less than 1 results in a compressive relationship such as that described by Fechner's law. The power function relationship between physical and perceptual scales has become known as *Stevens' power law*. It has been used to model many perceptions in color imaging, such as the prediction of lightness in the CIELAB color space.

Weber, Fechner, and Stevens formed the basis for many of the psychophysical techniques still used to develop and test color and appearance today. It is important to note the specific differences between Weber's goals and Fechner and Stevens' goals. In determining the amount of weight necessary to elicit a noticeable change in perceived weight, Weber was determining the threshold of detecting a change, or a just noticeable difference. Fechner and Stevens extended this to determine a scale of perceptual differences. Threshold and scaling techniques represent the main areas of psychophysical study for general color appearance.

2.5.1.1.1 Threshold techniques

These include detection, discrimination, and matching experiments. They are are designed to determine the perceptible limits to a change in a stimulus, or the just noticeable differences (JND). An example of a detection or discrimination technique used in imaging science is for developing and testing image compression algorithms. Two differing types of threshold JNDs can be calculated: ***absolute*** and ***difference***, which respectively determine the minimum amount and the smallest change detectable from a given stimulus. Three classical types of psychophysical techniques are used for determining thresholds: method of ***adjustment*** (where the observer must adjust the magnitude of the stimulus to reach a desired goal, or criterion[24]), method of ***limits*** (where the observer must report, either verbally or through a response-recording device, when an increasing stimulus is detected) and method of ***constant stimuli*** (as before, but now stimuli at various intensity levels around threshold are presented to the observer, in a random order). When the goal is to determine when two stimuli are not perceptible different, ***matching*** techniques (similar to the method of adjustment) are used instead.

[24] Example criterion might include adjusting a stimulus until it is just barely perceptible (for an absolute JND) or adjusting a stimulus until it is different from another (for a difference JND).

2.5.1.1.2 Scaling techniques

These are designed to produce a relationship between physical and perceptual magnitudes. Paired comparisons are effective for measuring very small differences between stimuli, but perceptual scaling techniques are needed to study larger differences efficiently. Stevens (1946) defined the properties of several types of measurement scales that are of utility in image quality assessment, including ordinal, interval, and ratio scales. There are three established psychometric methods which provide a one-dimensional scale of response differences: the method of ***rank order*** (which lets observers order the samples), the method of ***paired comparison*** (where observers are asked to choose between two stimuli based on some criterion), and the method of ***categories*** (which requires observers to sort stimuli into a limited number of categories; these usually have useful labels describing the attribute under study - e.g., excellent, very good, good, fair, poor, unsatisfactory-). Each of these scaling methods has proven to be of utility in image quality research. Unfortunately, the results of different rating experiments cannot readily be compared unless the scales are calibrated to some common standard, which has rarely been done.

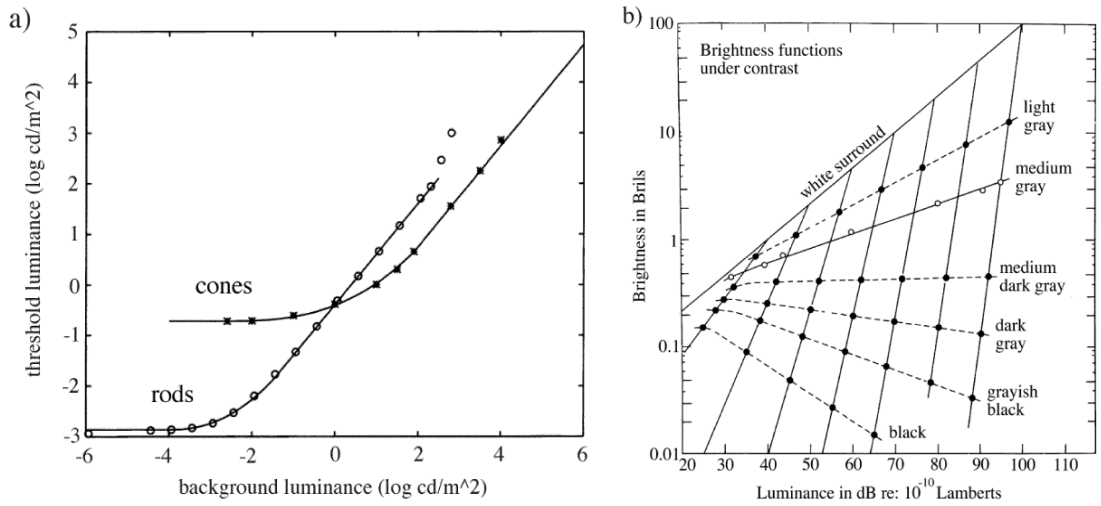


Figure 1: **Threshold and suprathreshold models of vision**: a) Threshold vs. intensity (TVI) functions for the rod and cone systems. The curves plot the smallest threshold increment ΔL necessary to see a spot against a uniform background with luminance L. b) Stevens' model of suprathreshold brightness and apparent contrast. The curves plot the changes in brightness and apparent contrast of gray targets and a white surround as the level of illumination rises (1 Bril = apparent brightness of a target with a luminance of 1μLambert). Reproduced from [148].

REFERENCES

Digital Image Processing

[17] BOVIK, A. C. *Handbook of Image and Video Processing.* 2nd ed. Elsevier Academic Press, 2010.

[18] GONZALEZ, R.C. and WOODS, R.E. *Digital Image Processing.* 2nd ed. Prentice-Hall, 2002.

[19] JÄHNE, B. *Digital Image Processing.* 5th ed. Springer-Verlag Berlin 2002.

[20] PRATT, W.K. *Digital Image Processing*. 3rd ed. John Wiley & Sons, 2001.

[21] RUSS, J.C. *The Image Processing Handbook.* 2nd ed. CRC Press, 1995.

[22] STOCKHAM, T. G., Jr. Image Processing in the Context of a Visual Model. *Proceedings of the IEEE*, Vol. 60, No. 7, July 1972.

[23] WINKLER G., *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods: A Mathematical Introduction*. 1995 - Springer-Verlag

[24] *Image Processing Fundamentals* (online course). Quantitative Imaging Group. Department of Imaging Science and Technology. Faculty of Applied Sciences. Delft University of Technology.Available at: http://www.ph.tn.tudelft.nl/Courses/FIP/frames/fip.html

Computacional vision

[25] BALLARD, Dana H. CM Brown. Computer Vision. *NY: Prentice Hill*, 1982.

[26] BARROW, Harry G.; TENENBAUM, Jay M. Computational vision. *Proceedings of the IEEE*, 1981, vol. 69, no 5, p. 572-595.

[27] FERWERDA, James A. Elements of early vision for computer graphics. *IEEE computer graphics and applications*, 2001, vol. 21, no 5, p. 22-33.

[28] KERSTEN, Daniel; MAMASSIAN, Pascal; YUILLE, Alan. Object perception as Bayesian inference. *Annual Rev. Psychol.*, 2004, vol. 55, p. 271-304.

[29] KNILL, David C.; RICHARDS, Whitman (ed.). *Perception as Bayesian inference*. Cambridge University Press, 1996.

[30] MARR, David; VISION, A. A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*, 1982, vol. 1, no 2.

[31] MARROQUIN, Jose; MITTER, Sanjoy; POGGIO, Tomaso. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American statistical association*, 1987, vol. 82, no 397, p. 76-89.

[32] PURVES, Dale; LOTTO, R. Beau. *Why we see what we do: An empirical theory of vision*. Sinauer Associates, 2003.

[33] WANDELL, B.A. *Foundations of Vision*. Stanford University. Sinauer Associates, 1995.

[34] YANTIS, S. *Visual perception: Essential Readings*. Taylor & Francis Group, 2001.

[35] www.webvision.com

Image Modeling

[36] ANDREWS, David F.; MALLOWS, Colin L. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1974, p. 99-102.

[37] MUMFORD, David; SHAH, Jayant. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 1989, vol. 42, no 5, p. 577-685.

[38] PORTILLA, Javier, et al. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image processing*, 2003, vol. 12, no 11, p. 1338-1351.

[39] RUDIN, Leonid I.; OSHER, Stanley; FATEMI, Emad. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 1992, vol. 60, no 1-4, p. 259-268.

[40] SIMONCELLI, Eero P.; OLSHAUSEN, Bruno A. Natural image statistics and neural representation. *Annual review of neuroscience*, 2001, vol. 24, no 1, p. 1193-1216.

Others

[41] LAY, D.C. *Linear Algebra and Its Applications*. Addison-Wesley Longman, Inc., Reading, Massachusetts, E.U.A., 1997.

[42] PAPOULIS, A. *Probability, Random Variables, and Stochastic Processes*. MacGraw-Hill, New York, incl. ed. 1991.

[43] JANSSEN, R. *Computational Image Quality*. SPIE Press, 2001.

Chapter 3

IMAGE QUALITY: ASSESSMENT AND IMPROVEMENT

Beauty in things exists in the mind which contemplates them.
-David Hume

Image capture, storage, transmission, transformation and display systems involve tradeoffs between system resources and output quality such as spatial and temporal resolution versus size and signal to noise ratio, speed versus accuracy, or luminance range versus gamut. How technology variables relate to customer quality preference has been for a long time the central question in image quality research. Although inherently involved in every imaging task, e.g. painting, the origins of image quality assessment are typically attributed to the invention of the earliest optical instruments, the telescope and microscope (1600-1620), and really gained attention with the introduction of photography (1860-1930), the development of television (1935-1955) and continue with digital imaging to the present day. During this time, a traditional device-dependent paradigm has been followed, regarding the quality of an image as the quality of the imaging system that generates it (i.e. perfection in its ability to capture the

visual world with as much detail and fidelity as the available technology allowed in each moment). Quality assessment thus reduced to a physical device characterization, while quality improvement concentrated on evolving the imaging technology (e.g. quality of the lens and the photosensitive substrate).

First vision models introduced in 1970s allowed to describe image quality not in physical but in perceptual terms. Up to now, several techniques have been proposed for quality assessment (see [68] for a review), but little effort has been done however to well state and understand the problem in terms of its main components, *quality* and *images*, define the former in relation to what are the latter used for and what requirements these uses impose on them [53]. Our understanding of the issue still remains very limited due to the complexity of the visuo-cognitive processes that underlie quality evaluation.

As a result, we are rather far from building a universal quality model and subjective experiments are to date the only widely recognized method of quantifying the actual perceived quality [69]. Moreover, it is regarded in [44] as the only "correct" one, given that humans are typically the "customers" of images. However, these methods are complex and expensive, and obtained results cannot be easily generalized and integrated in automatic systems, what has directed research activity towards the development of instrumental measures that, while substituting the human being, well correlate with subjective tests.[25] These measures find numerous applications in dynamically monitoring in intelligent networks and in optimization, parameter selection and benchmarking of image processing systems [44].

Within this context, the present chapter describes and analyzes the motivation, general ideas, and specific algorithms underlying the most representative image quality assessment methods available up to now, putting special stress on their interrelation. Only objective quality metrics are considered within this work. An in-depth discussion on subjective metrics can be found in [69] . The chapter is organized as follows. Section 3.1 introduces the basic concepts and necessary terminology within a classical framework. Section 3.2 presents a traditional classification of bottom-up approaches, where image quality is derived in terms of fidelity at different visual processing levels, ending with a discussion of the main limitations. Section 3.3 describes the new paradigms in image quality modeling during this decade, characterized by a top-down approach, where the hypothesized functionality of the HVS is modeled. Together with Section 3.4, they will provide the appropriate background and motivation to present the working definition of image quality followed by this thesis and its approach to image quality improvement.

[25] ***Subjective*** and ***objective*** adjectives have been therefore widely used to refer to the two main approaches to image quality assessment. In this context, the term ***objective*** means that no human interaction is required to derive these measures. The term ***computational*** has also been proposed to underline this fact.

3.1 The IQC framework for Image Quality Modeling

Attempts to provide solid foundations in image quality research go back to Burtleson's work in 1982, who highlights three main tasks: *(1)* identify perceptual attributes of quality, *(2)* determine how their scale values correlate with objective (in an instrumental sense) measures and *(3)* combine attribute values to predict overall quality [44]. It took, however, till 1989 to be formalized at the IS&T Annual Meeting as the proposed framework for image quality modeling, in order to decompose the task of relating customer quality preference to technology variables into more specific steps.

In a description given by Engeldrum in [49], the so-called ***Image Quality Circle*** (IQC) defines the following prerequisites:

- **Customer quality preference**: overall image quality rating as judged by customers in a specific situation.
- **Customer perceptions**: perceptual attributes of image quality (e.g. graininess, brightness, sharpness), called "***nesses***" to emphasize its perceptual nature.
- **Physical image parameters**: quantitative physical functions and parameters normally ascribed to image quality (e.g. modulation transfer, spectra, density, colour).
- **Technology variables***:* elements or parameters of the imaging system that are manipulated to change image quality (e.g. resolution, size, compression, printing method)

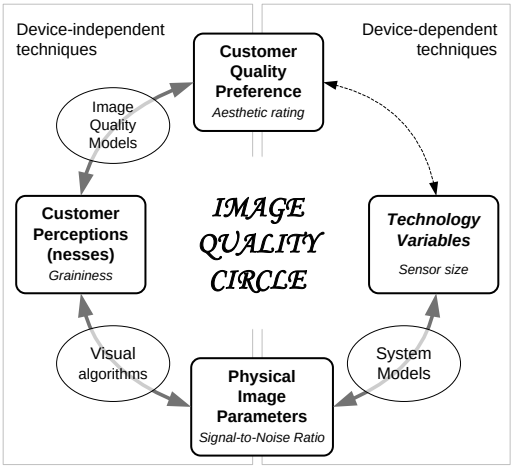


Figure 3.1. The Image Quality Circle (ICQ). An example from photography for typical prerequisites is shown in italics.

To describe how customer perceptions, physical parameters and technology variables are related, three more components were introduced:

- **Image quality models**[26]: empirical relation between perceptual attributes and customer quality rating.
- **Visual algorithms**: computation of the value of a perceptual attribute from a physical image parameter.
- **System models**: physical image parameters derivation from the technology variables.

[26] Also referred to as ***integration model*** and ***combination***, ***composition*** or ***integration rule***, terms borrowed from psychology literature and which may better capture the multidimensional nature of image quality. ***Pooling***, borrowed from statistics, is also common [68].

Depending on whether their combination is perceived as a new dimension or not, attributes of image quality are respectively said to be ***integral*** or ***separable***.[27] For example, while colour is an integral attribute, separable attributes such as graininess, sharpness and *"tone reproduction-ness"* (which are well-known nesses in photography[28]), guarantee an orthogonal representation in a ***ness space*** and propitiate the choice of a distance measure as integration rule. Generalized Mean Hypothesis and Minkowsky norm[29] are the preferred mathematical formalisms because of both, flexibility and good correlation with subjective tests.

All l_p norms are valid distance metrics in $\mathbf{R}^N$, which satisfy the following convenient conditions, and allow for consistent, direct interpretations of similarity: *i)* nonnegativity, $d_p(\mathbf{x},\mathbf{y}) \geq 0$; *ii)* identity, $d_p(\mathbf{x},\mathbf{y})$ 0 if and only if $\mathbf{x} = \mathbf{y}$; *iii)* **s**ymmetry: $d_p(\mathbf{x},\mathbf{y}) = d_p(\mathbf{y},\mathbf{x})$; and *iv)* triangular inequality: $d_p(\mathbf{x},\mathbf{z}) \leq d_p(\mathbf{x},\mathbf{y}) + d_p(\mathbf{y},\mathbf{z})$. In particular, the $p = 2$ case (proportional to the square root of the MSE) is the ordinary distance metric in N-dimensional Euclidean space.

$M_p(x_1, \ldots, x_n)$ is equivalent to: $\min\{x_i\}$ (for $p \to -\infty$); *harmonic*, *geometric*, *arithmetic*, or *quadratic* mean$\{x_i\}$ (for $p = -1, 0, 1$ *and* 2, *respectively*); and $\max\{x_i\}$ (for $p \to \infty$). Note also that $M_{-\infty} < M_{-1} < M_0 < M_1 < M_2 < M_\infty$

$$\textbf{\textit{Generalized mean:}} \qquad M_p(x_1,\ldots,x_n) = \left\{ \frac{1}{n} \sum_{i=1}^{n} x_i^p \right\}^{1/p} \qquad (3.1)$$

Image quality techniques that relate technology variables directly to customer quality preference are regarded by Johnson and Fairchild to be ***device-dependent***, in opposition to ***device-independent*** ones, which relate physical image parameters to customer quality preference [55]. The former group would belong to what they call the ***systems approach***, more related to psychophysics and statistics; while the latter, the ***fundamental approach***, would be related to psychophysics and vision modeling. Note, however, that these terms are not explicitly stated in the original formulation of the image quality circle.

[27] Analysis of preference judgements enable to determine whether we have isolated a single

[28] Although raised interest because of important applications, coding and compression artifacts are nesses quite less understood.

[29] ***Power mean, generalized mean, Hölder mean, mean of degree or order or power p***, all of the are equivalent [69].

3.1.1 Image Fidelity vs. Image Quality

There are two fundamentally different ways to consistently relate physical image parameters to human evaluation [65]: the ***impairment approach*** and the ***quality approach***. The former looks at decreases in image quality respect to some reference or ideal, while the latter attempts to model quality evaluation directly, independent of the reference. The impairment approach has traditionally been followed, reducing the problem of QA to measuring in a perceptual meaningful way the distance between a given image and a reference one, which is in turn considered to have "perfect" quality.

Although this may be the easiest way of modeling image quality, it presents two main drawbacks: first, fidelity and quality are not necessary synonymous. As noted by several authors (see for example [69], [68]), sharp and colourful images with high contrast are usually preferred. Second, a reference with which to compare is not available in most cases. If an ideal is defined at some abstraction level to be considered as a reference, then we should also consider terms such as *naturalness* and *realism*, for which Ferweda has described three different types within the context of computer graphics in [52]. Furthermore, aspects such as usefulness of the image, display type and properties, viewing conditions and image appealing also influence the customer preference[30].

This diversity has derived in three different ways of looking at the objective QA problem: on one hand, classical pixel-based, psychophysical and arbitrary criteria fidelity metrics belong to the so called ***error sensitivity*** paradigm; on the other, the ***structural similarity*** and the ***statistical*** or ***information fidelity*** paradigms have been recently introduced as an alternative to error sensitivity approaches.

Recent efforts have been made to develop more integrated approaches based on a general, perceptually relevant framework. Particularly active has been the research in ***visual appearance modeling***, which refers to *the prediction of the appearance of an image, or the difference in appearance of two images, accounting for as many known properties of the human visual system as possible, including those associated with neural processing* [56]. Unfortunately, our knowledge about the human perceptual system is very limited and most existing computational models follow a bottom-up approach based on rather simplistic stimuli. These models are most likely to be successful with artifacts (which are generally detrimental if detected), and near threshold (where visual phenomena are best understood, and simple difference measures are most likely to prove predictive). However, they are challenged by preferential attributes because images that differ substantially in appearance from variation in such attributes may, nonetheless, have equal perceived quality, which is the main reason why we prefer to follow a top-down approach in this thesis.

[30] See the work done by Fernandez, Fairchild and Braun in [51] for an example of analysis of observer and cultural variability.

3.1.2 Classification of Image Quality attributes

In [56], attributes contributing to perceived image quality, defined in a broad sense, are classified according to their **nature** (personal, aesthetic, artefactual, or preferential), which affects its amenability to objective description.

Although the objections to objective IQ assessment raised by a skeptic have some validity, they focus only on a subset of the attributes (personal –e.g. preserving a cherished memory or conveying a subject's essence- and aesthetic –e.g. lighting quality or composition-) that influence image quality. To obtain a more balanced perspective, two other types of attributes, *artefactual* and *preferential*, must also be considered .

3.1.2.1 Artefactual attributes

Artifacts are image features that, not being present in the original image or represented scene, appear in an image, often in the form of defects introduced by the imaging system and that nearly always lead to a loss of image quality when they are detected by an observer. Examples of such problems include *blurriness*, *noisiness*, *blocking* and *ringing compression artifacts* in JPEG coding. Assuming that an objective metric can be defined that is positively correlated with a given artefactual attribute, its impact on image quality can be adequately quantified provided that the *threshold point* (below which the attribute is not readily detectable by the HVS), and the *rate of quality loss* above threshold (where quality monotonically decreases with increasing values of the metric) can be characterized. Note that dependencies on scene content and observer sensitivity can be described in a statistical sense by characterizing the distributions of variations [56].

3.1.2.2 Preferential attributes

Unlike artefactual attributes, which degrade quality when detected, preferential attributes such as *contrast*, *colour balance*, *colourfulness* –saturation- or *memory colour reproduction*, are essentially always visible and have a range of preferred degrees depending upon both the tastes of the observer and the content of the scene. For example, while a low contrast is preferred in reproducing images taken on a bright, sunny day in order to lighten the deep shadows and make visible the detail within, images taken on an overcast day with flat lighting may be preferred at the higher contrast position, which causes it to appear crisper. Similarly, some observers may tend, on average, to prefer the "snappy", eye-catching appearance of higher contrast prints, whereas others may favor muted and understated rendition.

A probability density function that quantifies the relatively frequency of preference of different degrees of an attribute for some set of observers and scenes, frequently referred to as *preference distribution*, together with a *quality loss function*, which quantifies how rapidly quality falls off with the distance from the scene- and observer-specific optimum, are the principal tools for characterizing the impact of preferential attributes on image quality [65].

3.1.3 Threshold visibility and Suprathreshold judgments: appearance

Subjective image quality (SIQ), evaluated through judgment by human observers, has largely been dominated by error sensitivity, i.e. the probability of detecting an artifact. This corresponds to colour differences at the so called ***threshold level*** of perceptibility**,** often also referred to as ***just-noticeable differences*** (***JND***).[31]

It is important to make a distinction between *suprathreshold* judgements of image quality (such as preference judgements) and *threshold* visibility judgements (such as the visibility of image distortions). It is well accepted that JND predictions and suprathreshold appearance differences are not linearly related.[32]

Recently, a new framework presented by Fairchild in as iCAM (***image Colour Appearance Model***) focus on suprathreshold differences (*"how large is the perceived difference?"*) instead of error visibility [50].

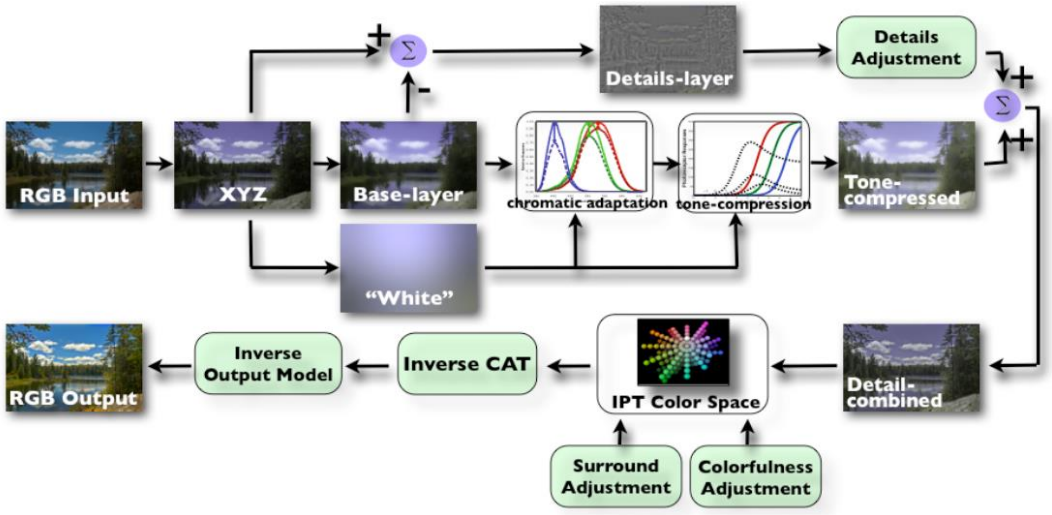


Figure 3.2. Flowchart of iCAM06 image appearance model. Based on the iCAM framework, incorporates the spatial processing models in the human visual system for contrast enhancement, photoreceptor light adaptation functions that enhance local details in highlights and shadows, and functions that predict a wide range of color appearance phenomena. Reproduced from [50].

---

[31] The field of psychophysics had its origins in the nineteenth century, when experimentation by Weber and others sought to relate discriminable differences in sensation to continuous physical properties such as weight. In 1860, Fechner proposed that such discriminable differences could be accumulated to form a quantitative scale of sensation, allowing perceptions to be mapped to numerical values.

[32] Paired comparisons are effective for measuring very small differences between stimuli, but perceptual scaling techniques are needed to study larger differences efficiently. Stevens (1946) defined the properties of several types of measurement scales that are of utility in image quality assessment, including *ordinal*, *interval*, and *ratio* scales. Such scales can in principle be obtained from various simple ranking tasks, including *rank ordering*, *categorical sorting*, and *magnitude estimation*. Each of these scaling methods has proven to be of utility in image quality research. Unfortunately, the results of different rating experiments cannot readily be compared unless the scales are calibrated to some common standard, which has rarely been done.

3.1.4 Metric performance evaluation

Comparison with subjective ratings is the only reliable method to evaluate the prediction performance of a quality metric with respect to some criteria that can be quantified with mathematical tools such as regression analysis. As a result, prediction performance of objective quality metrics is necessarily bounded by observers' agreement on the quality of the test set. This can only be described statistically, averaging over the opinions of a sufficiently large number of them. Results in the range of 90-95% have been obtained, what provides a quantitative upper limit on prediction performance. As a comparison, best-performing metrics achieve correlations around 80-85%, while the PSNR performance is just about 70%.[33]

A common previous step in metric evaluation is fitting objective and subjective scores, typically using logistic functions. The quality assessment method is then tested according to its: [34]

- **Prediction accuracy**: The ability to predict the subjective score with low error, characterized by ***a)*** the correlation coefficient between the subjective and objective scores after variance-weighted and ***b)*** non-linear regression analysis.
- **Prediction monotonicity**: The ability to accurately predict relative magnitudes of subjective scores, characterized by the Spearman rank-order correlation coefficient between the objective and subjective scores.
- **Prediction consistency**: The robustness of the predictor in assigning accurate scores over a range of different images, characterized by the outlier ratio.

3.1.5 Subjective Quality Assessment

Psychophysical scaling tools to measure subjective image quality have been available only for the last 25 to 35 years. Several subjective assessment methods covering different areas of service quality have been recommended and standardized by the International Telecommunications Union (ITU). Subjective testing for visual quality assessment has been formalized in ITU-R Rec. BT.500-11 (2002) and ITU-T Rec. P.910 (1999), which suggest standard viewing conditions, criteria for the selection of observers and test material, assessment procedures, and data analysis methods. The former was written with television applications in mind, whereas the latter is intended for multimedia applications.

The Mean Opinion Score (***MOS***) for each image, standardized in ITU-98, is computed as the mean of the Z-scores for that image, after removing any outliers.

---

[33] Data published by the Video Quality Experts Group (VQEG). Formed in 1997, it represents the most ambitious and comprehensive effort for performance evaluation of video quality assessment systems. The results of the first phase (1997-2000) concluded that the prediction performance of most evaluated models (included the PSNR) were statistically equivalent. For more details and results of the second phase (2003), refer to [73].

[34] As defined in by the VQEG.

3.2 Classical Objective Metric (Classification)

According to the availability of a reference or ideal, which is considered to have "perfect quality", quality metrics are commonly divided in [68][69]:

- **Full-reference (FR) metrics**: also referred to as ***fidelity*** or ***difference metrics***. Emphasize reduction on image quality as a deviation (***difference***, ***error***, ***degradation*** or ***impairment***) from a ***reference image***, i.e. loss of similarity. Thus, they do not consider the image itself, but a difference image –normally computed in a perceptually uniform space-. They are commonly used to evaluate the perceived strength of a degradation process such as lossy compression or coding, transmission, watermarking or rendering. They require access to the original image –reference-, which is an important restriction that severely limits the scope of application. On the other hand, human vision rarely requires a reference to determine visual quality. Nevertheless, most of image quality assessment methods proposed in the literature fall by far within this type.
- **No-reference (NR) metrics**: also referred to as ***blind quality assessment***. Emphasize image quality directly, not the difference. This is regarded as much a difficult task. The difficulty lies in telling apart distortions from actual content, i.e. they require interpreting or making some assumptions about the content, at least at a level of modeling some regularities or structure. Proposed methods typically consider the presence of very specific distortions types and are not easily generalized. They typically use prior knowledge about the distortions or artifacts introduced to estimate their strength, e.g. *blockiness* (which is the most prominent artifact of block-DCT based compression methods) and *blurriness* (they assume that the original, undistorted or ideal image contains sharp edges).
- **Reduced-reference (RR) metrics**: this group represents some compromise between the above two extremes. They use an available set of typically low-level features, which are supposed to have influence on perceived quality, as a reference to which compare.

Even if their prediction performance may not be as good, NR and RR metrics are regarded to be much more versatile and powerful than FR metrics, not only because the fewer restrictions imposed on the availability of a reference, but also the possibility to account for aspects such as image appealing attributes like sharpness or colourfulness [68].

Traditionally, fidelity metrics have followed three different approaches: 1) the mathematical or pixel-based comparison; 2) error sensitivity by means of human visual system modeling and 3) arbitrary criteria. These form the foundations of image quality assessment and are discussed in what follows. Section 3.3 presents two new paradigms on FR quality assessment that are closely related to this thesis' approach. For NR and RR approaches, see section 3.4.

3.2.1 Mathematical or Pixel-Based Fidelity Metrics

The goal of a signal fidelity measure is to compare two signals, x and y, by providing a quantitative score that describes the degree of fidelity or, conversely, the level of error, distortion, impairment or artifact between them. If one of the signals, e.g. x, is an original signal of acceptable (or perfect) quality, and the other, e.g. y, is a distorted version of it whose quality is being evaluated, then such a score may also be regarded as a measure of signal quality.

The simplest and more popular FR fidelity metrics are the so-called ***pixel-based metrics***. These are based on statistical measures of the magnitude of a pixel-by-pixel comparison $\boldsymbol{e_i = x_i - y_i}$. The most popular choice is the Mean Squared Error (***MSE***). For images of size $\boldsymbol{M}$ *x* $\boldsymbol{N}$ pixels and $\boldsymbol{B}$ colour bands:

$$MSE = \frac{1}{MNB}\sum_m\sum_n\sum_b\left[e(m,n,b)\right]^2 = {e_{RMS}}^2 = {\mu_e}^2 + {\sigma_e}^2$$

To have a measure that is comparable between different images, the MSE *difference* measure is often normalized respect to the power of the image, resulting in the ***SNR*** (Signal-to-Noise Ratio) and ***PSNR*** (Peak SNR) *fidelity* measures, where I is the maximum intensity value that a pixel can take (e.g., 255 for 8-bit images).

$$SNR = 10\cdot\log_{10}\left(\frac{\sum_i {(x)_i}^2}{\sum_i {(e)_i}^2}\right)\ (\text{dB}) \qquad PSNR = 10\cdot\log_{10}\left(\frac{I^2}{MSE}\right)\ (\text{dB})$$

MSE and related metrics based on the l_2 norm have been extensively used throughout the literature of image processing, communication, and many other signal processing fields because they are simple (parameter free and inexpensive to compute) and often the most convenient error measures for the purpose of algorithm optimization due to the very satisfying properties of *convexity*, *symmetry*, and *differentiability*. When combined with the tools of linear algebra, closed-form solutions can often be found for real problems.[35] In addition, as a measure of the energy of the error signal, it is preserved after linear orthogonal (or unitary) transforms such as the Fourier transform (Parseval's theorem). This property distinguishes d_2 from the other l_p energy measures, which are not energy preserving.

Nevertheless, the MSE has long been critized for its poor correlation with perceived image quality. This becomes clear in Figure 3.3.

[35] Minimum-MSE (i.e. MMSE, equivalent to Maximum Likelihood Estimation - MLE - for independent measurement errors with normal distribution) optimization problems often have closed-form analytical solutions, and when they don´t, iterative numerical optimization procedures are often easy to formulate, since the gradient and the Hessian matrix of the MSE are easy to compute.

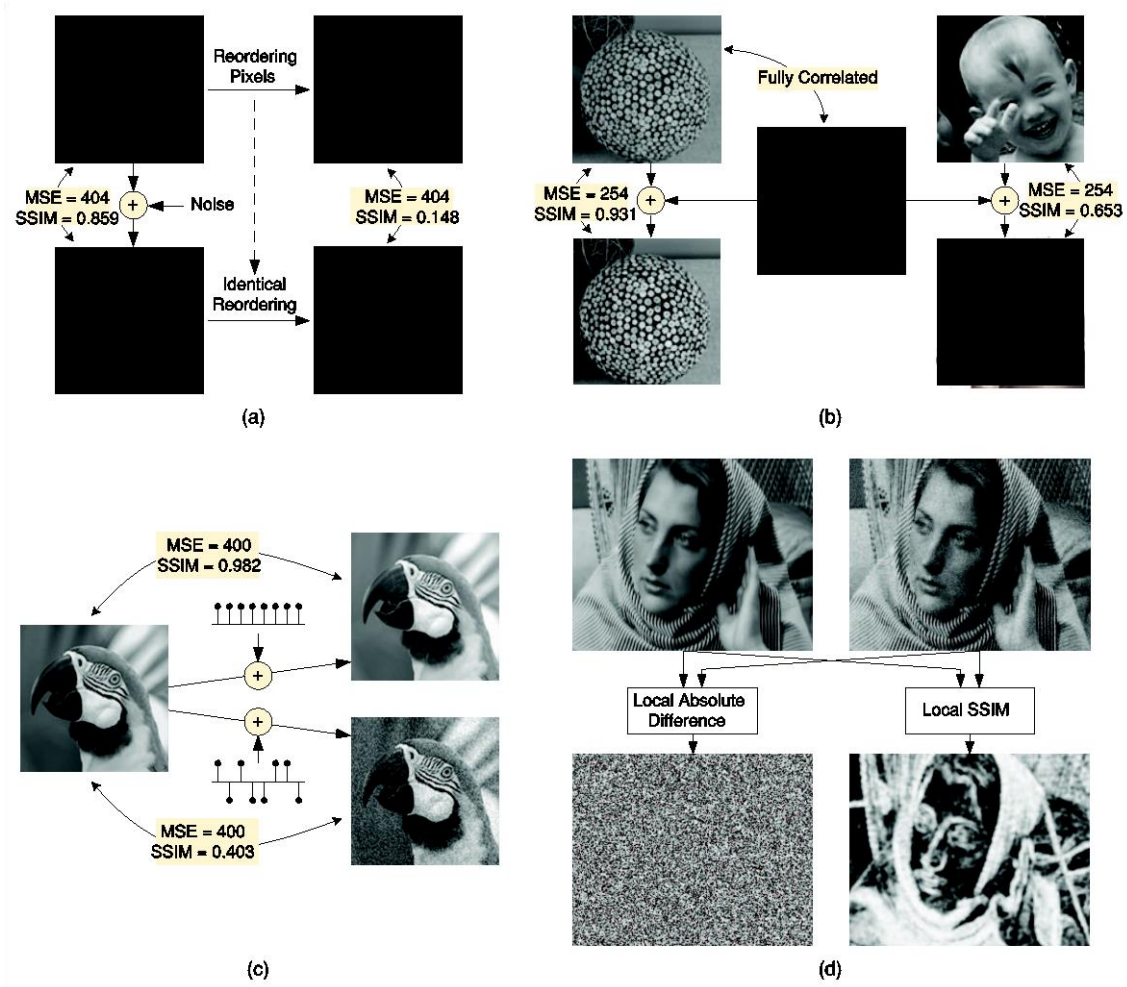


Figure 3.3. Failures of the MSE and other *lp* metrics to predict perceived differences (see explanation in the text below). Reproduced from [65].

A direct explanation of the apparent failure of the MSE (and in general any Minkowsky) metric in these examples is based on the fact that, whenever one chooses to use an l_p norm to predict perceptual image quality, a number of questionable assumptions have been made [65]:

- *Image fidelity is independent of any spatial relationships between image signal samples*. According to this, changing the spatial ordering of the image signal samples should not affect the distortion measurement.

- *Image fidelity is independent of any relationships between the image signal and error signal*. I.e., for the same error signal, no matter what the underlying image signal is, the distortion measurement remains the same.

- *Perceptual image quality is determined by the magnitude of the error signal only*. As a result, changing the signs of the error signal samples has no effect on the distortion measurement.

- *All signal samples are of equal importance in perceptual image quality*.

Unfortunately, none of these assumptions holds (even roughly) for perceptual image quality assessment, as demonstrated in Figure 3.3:

- In ***(a)***, an original image (top left) is distorted by adding independent white Gaussian noise (bottom left). In the top-right image, the pixels are reordered by sorting pixel intensity values. The same reordering process is applied to the bottom-left image to create the bottom-right image. The MSE (and any l_p metric) between the two left images and between the two right images are the same, but the bottom-right image appears much noisier than the bottom-left image.

- In ***(b)***, two original images (top left and top right) are distorted by adding the same error image (middle), which is fully correlated with the top-left image. The MSE (and any l_p metric) between the two left images and between the two right images are the same, but the perceived distortion of the bottom-right image is much stronger than that of the bottom-left image.

- In ***(c)***, an original image (left) is distorted by adding a positive constant (top right) and by adding the same constant, but with random signs (bottom right). The MSE (or any l_p metric) between the original and any of the right images are the same, but the right images exhibit drastically different visual distortions.

- In ***(d)***, an original image (top left) is distorted by adding independent white Gaussian noise (top right). The energy distribution of the absolute difference signal (bottom left, enhanced for visibility), which is the basis in computing all l_p metric, is uniform. However, the perceived noise level is space variant, which is reflected in the SSIM map (bottom right, enhanced for visibility) [67].

One potential solution to overcome the first problem is to apply, prior to the Minkowski metric, an image transform T ideally characterized by: *i)* decoupling (or at least decorrelation) of image samples, *ii)* preservation of visual information[36], and *iii)* reduction of dimensionality. Since such a transform can decouple the dependencies between image signal samples without losing important visual information, one may say that the "*structure*" of the image signal is well captured by the transform domain representation.

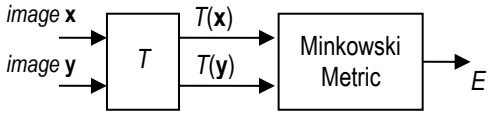


Figure 3.4. An image transform prior to a Minkowsky metric may potentially reduce the dependencies between signal samples, thus improving an image quality metric.

The MSE is only suitable as a fidelity metric for direct application in image space in those cases where the distortion has zero mean and is independently

[36] Presumably, an inverse transform that can reconstruct the image signals in the spatial domain should exist.

distributed, yielding estimates that better correlate with perceived distortion. This approximation specially holds for graininess, normally assumed to have AWGN (Additive White Gaussian Noise) properties. This fact has been modeled in approaches based on the human visual system (***HVS***) through a set of psychophysical features of human vision such as the *contrast sensitivity function* and *masking effect*, which are introduced in next section.

3.2.2 Psychophysical Fidelity Metrics

For FR QA methods, modeling the HVS has been regarded as the most suitable paradigm for achieving better quality predictions, for which most recent models are based on multiscale, bandpass and oriented linear transformations (like ***wavelet*** ones, see Chapter 2, *Natural Scene Statistics*).

3.2.2.1 Single-scale metrics

First psychophysical metrics used ***single-channel linear models***, regarding the visual system as a single spatial filter. Quality measures based on linear HVS models assess image quality in three steps. First, an error image is computed as the difference between the original image and the restored image. Second, the error image is weighted by a frequency response of the HVS given by a low-pass contrast sensitivity function (CSF). Finally, a signal-to-noise ratio is computed [48][55][66].

These quality measures can take into account the effects of image dimensions, viewing distance, printing resolution, and ambient illumination. According to [69], these are between the first efforts to recognize the importance of applying vision science to image processing.

Further introduction of logarithmic nonlinearity to compute cone responses finally resulted in S-CIELAB (spatial extension for CIELAB) and ST-CIELAB (temporal extension for S-CIELAB) image appearance models [70][72].

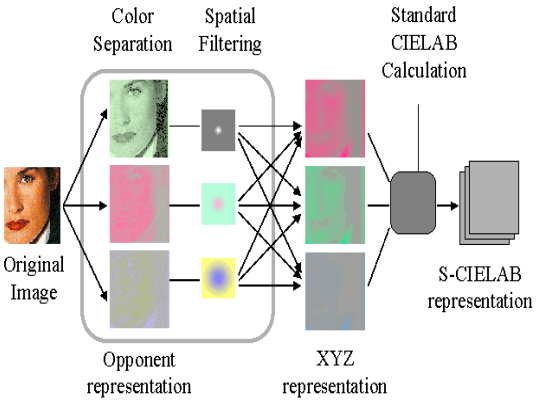


Figure 3.5 S-CIELAB, a spatial extension to CIELAB Model. Adapted from [72].

3.2.2.2 Multiscale metrics

Modeling pattern adaptation and masking effects required the introduction of ***multi-channel models***, which divide spatial frequencies into channels with different sensitivity. Well known is the Visual Differences Predictor (VDP), in which the amplitudes at different channels are (1) non-linearly transformed to account for adaptation processes, (2) thresholded, (3) converted to detection probabilities and (4) combined to produce a map of visible differences (i.e. the probability of detecting an artifact at a given image position).

The underlying premise of these metrics is that the sensitivities of the HVS are different for different aspects of the visual signal that it perceives, such as brightness, contrast, frequency content, and the interaction between different signal components, and it makes sense to compute the strength of the error between the test and the reference signals once the different sensitivities of the HVS have been taken into account [68].

Despite being more or less complex, all of these metrics share a quite similar architecture, shown in Figure 3.6, where several functional properties of early stages of the HVS are regarded as independent sequential processes with the aim of weighting different aspects of the error image according to their visibility, previously determined by psychophysical and physiological experiments.

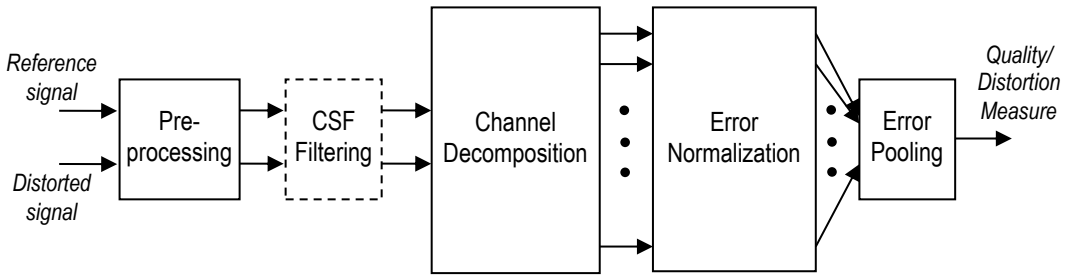


Figure 3.6. General modular framework for full reference quality assessment based on psychophysical error sensitivity. Note that the CSF feature can be implemented either as a separate stage (as shown) or within *Error Normalization* block. Reproduced from [65].

Even if they differ in description detail and implementation issues, the stages represented in Figure 3.6 typically correspond to the following conceptual steps [44][68][69]:

1. **Colour preprocessing**: simulation of display characteristics (by computing luminance levels from pixel values) and eye optics (the most important aspect is the point spread function), followed by a transformation to a proper colour space such as CIELAB in order to account for ***opponent colour encoding*** and ***lightness non-linearity***.
2. **CSF filtering**: the ***contrast sensitivity function*** (CSF) accounts for the variation of visual sensitivity as a function of spatial (and maybe also temporal) frequency. It typically exhibits a band-pass behavior for luminance channels, and low-pass for the chromatic ones [37].
3. **Multi-channel decomposition**: images are separated into sub-bands or *channels* and selectively processed at different scales of both, spatial and temporal information. DCT and wavelet are preferred over Fourier

[37] The HVS is a nonlinear, spatially varying system. A measure of the nonlinear HVS response to a single frequency, called the *contrast threshold function* (CTF), is given by the minimum amplitude necessary to just detect a sine wave of a given angular spatial frequency [48]. Inverting a CTF gives a frequency response, called the *contrast sensitivity function* (CSF), which is a linear spatially invariant approximation to the HVS.

transforms. [38]

4. **Lightness adaptation**: as described by the Weber-Fechner law, lightness adaptation mechanisms mostly disregard absolute luminance values, considering only local variations in relation to a surrounding background. This is normally referred as ***local contrast***.
5. **Error normalization**: visibility threshold at each point is computed based on local characteristics of the image and used to weight channel error contributions and account for ***masking effects*** (i.e. when the visibility of a stimulus is conditioned by the presence of another, e.g. *contrast, edge, texture masking*).
6. **Pooling**: rule for integrating information from several channels, sometimes performed at different levels and normally with the purpose of obtaining a single rating of quality. In some situations, however, an error *map* or image, which gives the quality score at each location, is desirable. The most commonly used pooling methods adopt an l_p norm.

3.2.3 Arbitrary Criteria Metrics or *the Engineering approach*

Although they may not be as versatile as metrics based on multi-channel vision models, extraction and analysis of certain image features or artifacts based on a priori knowledge of the degradation process generally allows more efficient implementations.

This is the main idea lying under the so call arbitrary criteria or engineering approach to objective image quality metric design, typically based on perceptual weighting of quantization and coding noise, artifact visibility (specially for DCT-based compression formats, such as jpeg and its derivates) [69].

While pixel- and HVS-based metrics typically belong to the full-reference class, the engineering approach is almost the only possibility for reduced and no-reference metrics. These are described in section 3.4.

The engineering approach, which has gain popularity in recent years, is based primarily on the extraction and analysis of certain features or artifacts in image, either structural elements such as edges, or specific distortions that are introduced at capture, compression or transmission. The metrics look for the strength of these features in the image to estimate overall quality. This does not necessarily mean that such metrics disregard human vision, as they often consider psychophysical effects as well, but image analysis rather than fundamental vision modeling is the conceptual basis for their design.

[38] It is well known that a large number of neurons in the primary visual cortex are tuned to visual stimuli with specific spatial locations, frequencies, and orientations [68].

3.2.4 Limitations

For FR QA methods, modeling the HVS has been regarded as the most suitable paradigm for achieving better quality predictions. Nevertheless, traditional HVS-based approaches also suffer from several widely recognized limitations, in great part derived from the extrapolation of results from psychophysical experiments to the complexity of natural scenes. In general, FR quality metrics based on error sensitivity have the following limitations.

- On one hand, the accuracy of the reproduction is clearly just a part of image quality assessment, and may not be necessarily correlated. Examples are image transformations that result in visible but not objectionable distortions, such as the application of a tone reproduction curve that allows a better recognition of fine detail, colour saturation, deblurring algorithms and, in general, every image enhancement technique.
- On the other hand, little effort has been done in answering the question "***how large is the perceived difference?***" [69]. In [50] it is described an approach recently proposed to answer this question in terms of a unified framework for image appearance, differences and quality. Developed fidelity metrics are often applied in cases where the distortion magnitude falls quite out of threshold levels. However, it is well accepted that just noticeable differences (***JND***) predictions and suprathreshold appearance differences are not linearly related [68].
- In order to get consistent results, psychophysical and physiological experiments use patterns such as uniform colour patches against a background, bars or sinusoidal gratings[39]. It is not clear whether results obtained with such relative simple patterns are applicable to real images, which exhibit quite a larger complexity which remains unexplored
- As introduced before, ness separability is desirable in order to be successfully combined through a Minkowsky metric or a generalized mean. However, channel decomposition is commonly performed through DCT and wavelet linear techniques, which result in intra– and inter- channel dependencies, especially when non-orthogonal over-complete decompositions such as the steerable pyramid are used for orientation selectivity and translation invariance purposes. These dependencies should be decoupled by means of more complex schemes based on decorrelation approaches, such as principal component analysis.

Finally, complex image analysis processes, such as segmentation, object recognition or image understanding, play, together with higher cognitive factors, an important role when assessing image quality. Prior information about image content, point of interest and provided instructions highly influence the result.

[39] As example, the CIE Lab Delta E metric is intended to be used on large uniform color targets (at least 2° visual angle in size) [72].

3.3 New Paradigms in FR Image Quality Modeling

Traditional approaches to image quality assessment presented in section 3.2 follow a bottom-up approach, where models of the HVS are used to derive quality metrics. The effectiveness of these methods depends on how much the HVS is understood and how accurately the simulation can be implemented. By contrast, recent approaches to quality assessment follow a top-down approach where the hypothesized functionality of the HVS is modeled. A top-down approach may lead to significantly simplified algorithms, but relies on the goodness of the underlying hypothesis.

The two new paradigms presented in this section, ***structural similarity*** and ***information fidelity***, represent first attempts to formulate image quality neither in terms of distortion visibility nor signal processing, but regarding the visuo-cognitive process as an essential information processing stage in human interaction. Because of their vital importance as motivating preliminary for next chapters, they are described in more detail than those approaches in previous section.

Finally, being still at a preliminary stage, they are both, easy to extend and very encouraging. This is because of a simpler formulation (what makes them more tractable than traditional methods in optimization tasks), designing freedom and lower computational complexity.

3.3.1 Structural Similarity

Natural images are highly structured and present strong dependencies among spatially proximate samples. Therefore, image samples define a vector space of much larger dimensionality than that of the subspace where natural images lay. If we view the human visual system as an ideal information extractor that seeks to identify and recognize objects in the visual scene, then it must be highly sensitive to the *structural* distortions (e.g., noise, blur, or lossy compression artifacts) and automatically compensates for the *nonstructural* distortions (e.g., a change of luminance or brightness, a change of contrast, or a spatial shift). Consequently, an effective objective signal fidelity measure should simulate this functionality [67]. This can intuitively explain why, although all the images in Figure 3.3 have the same MSE value, their respective distortions are perceived with significant different strength.

Consequently, approaches to quality assessment based on structural similarity attempt to measure structural similarity between reference and test image by mean of more complex metrics than the Minkowsky one. Although, these should nevertheless capture the properties of the avoided de-correlating transform $\boldsymbol{T}$ introduced in 3.2.2, they have the advantage of being simpler, computationally more efficient and do not depend on psychophysical modeling of the HVS. While this could lead to consider them among Arbitrary Criteria Metrics, the fact that they follow a substantially different design principle lend themselves to be considered apart.

Geometrical interpretation of Minkowsky metrics in the image vector space on one hand, and the perspective of image formation on the other, provides important insights for structural similarity metrics design. From the former, all the images with the same r^2 MSE with respect to a reference image lie on the same hyper-sphere of radius r, despite having quite different visual quality. What perceptually differentiates them is then not the distortion strength, but its type (i.e. geometrically, it is not the length, but the direction of the distortion vector).

From an image formation point of view, every pixel value corresponds to a captured luminance from the real-world, product of the illumination falling on the scene and the reflectance of the objects. Reflectance is an intrinsic property of the object that characterizes it and allows us its recognition. Because it does not depend on illuminance, it is reasonable to separate both components, illuminance and reflectance, as well as noise introduced by the image capture process, prior to computing the similarity between reference and test images.

In addition, recall that this concept is not new at all: *brightness* (or *tone-reproduction-ness*), *sharpness* and *graininess* were identified in 3.1 as well-known separable nesses in photography. As mentioned there, if properly characterized, these guarantee an orthogonal representation in a ness space and propitiate the choice of a distance measure as integration rule.

3.3.1.1 The Structural SIMilarity (SSIM) Index

A particular implementation of this idea is given by Wang, Bovik and Sheik in [66], where the method, first introduced in 2002 under the name of *Universal Quality Index*, is generalized and improved. Although it is here described because of its close relation to the philosophy employed in this thesis, note that many other approaches may emerge from the same concepts.

The basic form of SSIM is very easy to understand. Suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ are local image patches taken from the same location of two images that are being compared. The local SSIM index measures the similarities of three elements of the image patches: the similarity $l(\mathbf{x}, \mathbf{y})$ of the local patch *luminances* (brightness values), the similarity $c(\mathbf{x}, \mathbf{y})$ of the local patch *contrasts*, and the similarity $s(\mathbf{x}, \mathbf{y})$ of the local patch *structures*. These local similarities are expressed using simple, easily computed statistics, and combined together to form local SSIM.

Like many other inverse problems in vision, this component separation task is ill-posed. However, simple statistical estimates based on locality and homogeneity assumptions usually serve the purpose. The separation is defined as a projective transformation which maps the N-dimensional image vector space to a 3-dimensional space (luminance, contrast and structure components). This is shown Figure 3.7, where the image is represented by the image vector $\boldsymbol{x}$ in an N-dimensional space.[40]

[40] Here represented as a 3-dimensional space because of obvious reasons.

The SSIM system diagram is shown in Figure 3.7 below, where:

- ***Luminance*** measurement, μ_x, is computed as the mean intensity, i.e. a 0th order approximation. Geometrically, this is the image projection onto the $x_i = x_j = \ldots = x_k$ axis

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (3.2)$$

- ***Contrast*** measurement, σ_x, is obtained as the unbiased estimate of the standard intensity deviation. This is equivalent to the image projection onto the hyperplane $\sum x_i = 0$,

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)\right)^{1/2} \qquad (3.3)$$

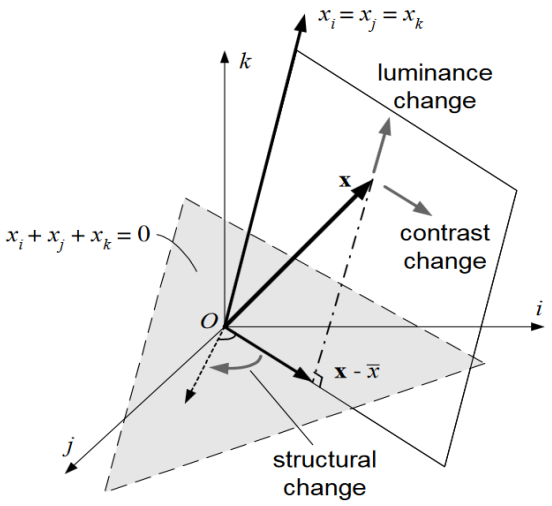


Figure 3.7 Separation of luminance, contrast and structural changes from a reference image *x* in the image space. This is an illustration in three-dimensional space. In practice, the number of dimensions is equal to the number of image pixels. Reproduced from [66].

- ***Image structure*** is determined by the remaining image after luminance subtraction and contrast normalization (division), $(x-\mu_x)/\sigma_x$, what geometrically represents the direction of the resulting vector lying in the $\sum x_i = 0$ hyperplane.

Note also that, according to the above definitions, an image distortion is described by a change in each of these three independent components.

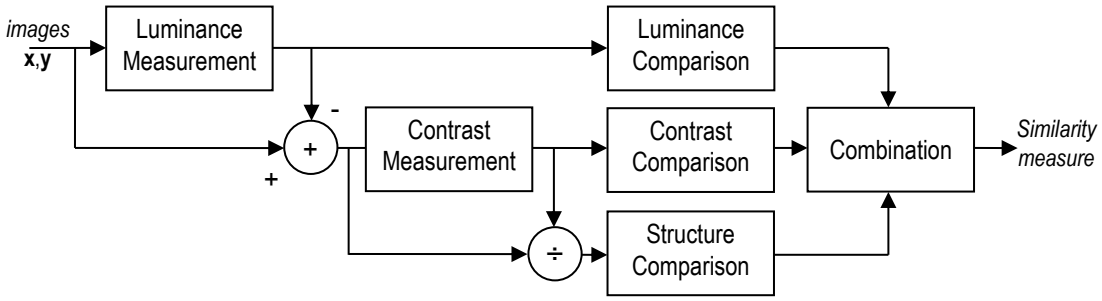


Figure 3.8 Diagram of image similarity measurement system. Adapted from [67].

Component comparisons between a reference image ***x*** and a test image ***y*** are then defined under symmetry ($S(x,y) = S(y,x)$), boundedness ($S(x,y) \le 1$) and unique-maximum ($S_{max}(x,y) = 1$) constrains as follows

- ***Luminance*** comparison

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^{\,2} + \mu_y^{\,2} + C_1} \qquad (3.4)$$

- ***Contrast*** comparison

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^{\,2} + \sigma_y^{\,2} + C_2} \qquad (3.5)$$

- ***Structure*** comparison

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \qquad (3.6) \qquad \sigma_{xy}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y) \qquad (3.7)$$

where, C_1, C_2 and C_3 are small constants to avoid instability. Finally, these comparisons are combined together in the Structural SIMilarity (SSIM) index as given by the following integration rule

$$SSIM(x,y)=[l(x,y)]^{\alpha}\cdot[c(x,y)]^{\beta}\cdot[s(x,y)]^{\gamma} \qquad (3.8)$$

Note that the SSIM index defines an error measure in a *locally adaptive*, non-linear, and input-dependent coordinate system.

Finally, because natural images statistical features are spatially non-stationary and, at the same time, some image distortions may also be space-variant, the authors suggest applying the SSIM index locally, what can provide a spatial quality map instead of a single score.

The SSIM index approach is very attractive not only because of its remarkable image quality prediction accuracy across a wide variety of image and distortion types, as can be seen in Figure 3.3, but also because of its simple formulation and low computational complexity, being much more tractable in optimization tasks. While taking a variety of forms, depending on whether implemented at a single or multiple scales, the best results are obtained when computed over a range of scales in the wavelet domain (called 'CW-SSIM'), what makes it simultaneously robust with respect to translations and changes in luminance and contrast [67].

3.3.2 Information Fidelity

This section presents two very recent related full-reference fidelity metrics based on theoretical approaches to image quality formulated in terms of *statistical information* rather than *signal fidelity* criteria.

Statistical modeling plays a fundamental role as *a priori* knowledge in source coding, estimation and decision problems, for which the information theory was developed, first by Shannon in 1948 within the context of signal transmission over communication channels. The fact that information acquisition, transmission, manipulation and storage are common processes present in almost every task has caused that information theory is nowadays used in a lot of different research and development fields.[41]

The fundamental hypothesis of the visuo-statistical approach to image processing and computer vision is that the biological visual system has evolved toward an efficient adaptation to the statistical properties of the visual environment. If this assumption is accepted, it is reasonable to further consider it in terms of *coding efficiency* in an information theory framework to link environment statistics and neural response. First suggestions on this idea go back

[41] In fact, the purpose of our senses is to provide our brain with a reliable communication with the environment and, thus, their study should also benefit from information theory.

more than 40 years ago to Attneave and Barlow's works in respectively, information theory and neurobiology. However, it is not till last decade that the development in computational tools and statistical modeling has enabled its empirical validation [40].

Information fidelity approaches to image quality assessment hypothesize that the statistical information shared between a test and a reference image, i.e. ***mutual information***, should relate well with perceived quality. Intuitively, mutual information measures how much knowing of the test image reduces the uncertainty about the original one, reducing to 0 for independent images or achieving a value of 1 for identical ones.

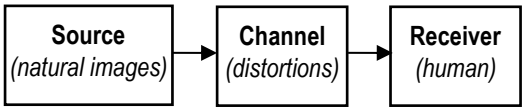


Figure 3.9 Information-fidelity approach. Image distortions are modelled as a result of the limited capacity of a communication channel that reduces the amount of information about the reference image that can be extracted from the test image.

Following this idea, reference and test images are respectively considered as the input and output of a communications channel (Figure 3.9) that limits the amount of information that can pass through it. This model accounts for reductions on visual information present in the reference image due to distortion processes such as compression, blurring or noise addition, imposing an upper limit on channel capacity, which is eventually determined by the image source statistics.

3.3.2.1 The Information Fidelity Criterion

Both information-based approaches presented here are described in the wavelet domain, where natural scenes are statistically described by a GSM model (see *'Statistical Image Modeling'* in Chapter 2 for a brief introduction on the convenience of this transform and a description of GSM models). To simplify, in what follows just one sub-band of the wavelet image decomposition is considered. The procedure is later generalized for multiple sub-bands, each of these being a GSM random field (***RF***), ***C***, product of two stationary and independent RFs, ***S*** and ***U***, as defined by eq. (3.9)

$$C = S \cdot U = \{S_i \cdot U_i : i \in I\} \qquad (3.9)$$

$$p_{C_i|S_i}(c_i|s_i) \approx N\left(0, s_i^2 \sigma_U^2\right) \qquad (3.10)$$

where S is a RF of positive scalars, ***U*** is a Gaussian scalar RF with zero mean and variance σ^2_U, and ***I*** denotes the set of spatial indices for the RF. Note that, conditioned on S_i, C_i are normally distributed accordingly to (3.10).

The distortion model for each sub-band is simply a signal attenuation ***G***, which captures changes in image contrast that result either from blur distortion or variations in lighting, and additive Gaussian noise ***V***

$$D = GC + V = \{g_i C_i + V_i : i \in I\} \qquad (3.11)$$

The authors argue in [64] that this gathers what the HVS perceives as natural distortions and what should have driven its evolution: ***blur*** *due to lens*, ***brightness and contrast*** *stretches due to changes in ambient lighting*, and *white* ***noise*** *due to*

photon noise or internal neuron noise[42]. Thus, modeling source and distortion in this way should be a dual approximation of HVS signal estimator modeling.

The Information Fidelity Criterion (***IFC***) is then defined as the statistical information that is shared between the source and the distorted images, i.e. mutual information, computed (3.12) and summed over each band (3.13), under the assumption that the source $\boldsymbol{S}^N$ is known ($\boldsymbol{S}^N = \boldsymbol{s}^N$)[43]

$$I\left(C^N; D^N \middle| S^N = s^N\right) = \sum_{i=1}^{N} I\left(C_i; D_i \middle| S_i = s_i\right) = \frac{1}{2}\sum_{i=1}^{N} \log_2\left(1 + \frac{g_i^{\,2} s_i^{\,2} \sigma_U^{\,2}}{\sigma_V^{\,2}}\right) \qquad \textbf{(3.12)}$$

$$IFC = \sum_{k \in subbands} I\left(C^{N_k,k}; D^{N_k,k} \middle| S^{N_k,k} = s^{N_k,k}\right) \qquad \textbf{(3.13)}$$

Remark at this point that: first, for orientation selectivity and translation invariance, non-orthogonal over-complete decompositions such as the steerable pyramid are commonly used, introducing correlations between coefficients, in which case eq. (3.12) no longer holds [44]. Second, despite the mathematically convenience (due to the nonlinear dependence among the $\boldsymbol{C}^N$ by way of $\boldsymbol{S}$) of considering a known source image, it has the drawback of reducing the application scope of the IFC to a FR metric. Third, as noted in [64], just one realization is available from the source and the distortion RFs, from which their statistical properties must be estimated (this requires the ergodicity assumption). Fourth, eq. (3.13) assumes that bands are independent, what is not proved by the authors.

3.3.2.2 The Visual Information Fidelity Measure

The described IFC assumes that the receiver (HVS) is able to extract all the information carried in an image, i.e. it does not consider the limitations of the HVS. Observing that these can also be modeled as a communications channel, Sheik and Bovik proposed in 2004 to obtain a relative rather than absolute measure of the mutual information [65]. The new situation is depicted in Figure 3.10, where mutual information between ***C*** and ***E*** quantifies the information that the brain could ideally extract from the reference image, whereas the mutual information between ***C*** and ***F*** quantifies the corresponding information that could be extracted from the test image.

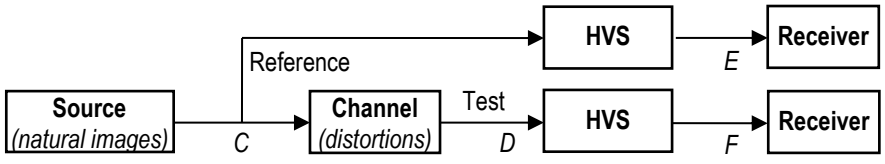


Figure 3.10. Block Diagram of the VIF Quality Asessment System. Reproduced from [65].

[42] Note that this is just an approximation. Neither photon noise nor neuron noise are white.

[43] $\boldsymbol{S}^N$ are the corresponding ***N*** elements of ***S*** and $\boldsymbol{s}^N$ denotes a realization of $\boldsymbol{S}^N$.

[44] Two approximations to solve this problem can be found in [64].

As mentioned earlier, information-based approaches consider that the HVS has evolved to better adapt to natural scene statistics, thus several aspects of the HVS are already described in the source model. Because of this, it is argued in [65] that enough improved performance over the IFC is obtained by using just an additive noise model for the HVS, i.e. $\boldsymbol{E=C+N}$, $\boldsymbol{F=D+N}$, where the noise field $\boldsymbol{N}$ is normally distributed with $\boldsymbol{\sigma_N^2}$ variance (covariance matrix is a multiple of the identity) and is assumed to be independent of $\boldsymbol{C}$

$$I\left(C^N;E^N\middle|S^N=s^N\right)=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\log_2\left(1+\frac{s_i^{\,2}\lambda_j}{\sigma_N^{\,2}}\right) \qquad \textbf{(3.14)}$$

$$I\left(D^N;F^N\middle|S^N=s^N\right)=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\log_2\left(1+\frac{g_i^{\,2}s_i^{\,2}\lambda_j}{\sigma_N^{\,2}+\sigma_V^{\,2}}\right) \qquad \textbf{(3.15)}$$

Note that eq. (3.15) is the same as eq. (3.12), with addition of new noise term in the denominator[45]. The Visual Information Fidelity (VIF) measure is given by

$$VIF=\frac{\sum_{j\in subbands}I\left(C^{N,j};F^{N,j}\middle|S^{N,j}=s^{N,j}\right)}{\sum_{j\in subbands}I\left(D^{N,j};F^{N,j}\middle|S^{N,j}=s^{N,j}\right)} \qquad \textbf{(3.16)}$$

The most important property of the VIF criterion compared to other quality metrics is that, unlike the IFC criterion –which is limited to the interval [0, 1]-, it can take values greater than 1. This translates in that the brain could extract more information from the distorted image than from the original one or, what is the same, it captures improvements in the image quality, which could be caused by contrast or detail enhancement, resulting in more information delivered to the brain by the HVS.

The importance of information theoretic approaches to image quality is twofold. First, they place a limit on the amount of information that hypothetically could be extracted from an image. Second, because modeling natural scenes and the human visual system is regarded as a dual problem, they represent the counterpart of HVS-based approaches. However, in contrast to HVS-based methods or signal fidelity measures presented in section 3.2, success of information fidelity approaches critically relies on an appropriate statistical characterization of the source and the channel instead of complex computational models developed to describe the results of psychophysical experiments. As a result, while the obtained fidelity criterions functionally capture the HVS sensitivities, they do not require neither parameters associated with display device, viewing configuration, etc. nor training data. Moreover, they outperform state of the art HVS-based and structural fidelity methods, accordingly to the authors.

[45] Except for $\boldsymbol{\lambda_j}$, eigenvalues of the covariance matrix $\boldsymbol{C_U}$. These could also be used in (3.12) where it has been assumed that $\boldsymbol{C_U=\sigma_U^2 I}$, though.

3.4 No-Reference Metrics

In contrast with fidelity metrics, which look at decreases in image quality in terms of similarity or closeness to some reference or ideal, no-reference (***NR***) metrics, also referred to as *single ended* or *blind quality assessment*, attempt to model quality evaluation directly, independently of a reference. This addresses a fundamental distinction between ***fidelity*** and ***quality*** [57].[46] NR metrics are required in many applications where a reference is not available, such as image interpolation, intelligent memory management, implementation of transparent and competitive ratings of quality of service (***QoS***) and quality of products, etc.

Reduced- and no-reference metrics almost exclusively follow the so-called *engineering approach*, which is conceptually based on image analysis rather than fundamental vision modeling. E. Cavides and F. Oberty identify in [47] three principles which allow measuring quality in a no-reference manner:

- Presence of distortions due to transmission (e.g., noise), compression (e.g., JPEG artifacts) and image processing (e.g., clipping), which have a monotonic although not continuous effect on quality, being the best the one of the distortion-free image.
- Enhanced images show improved attributes such as sharpness and contrast, in addition to reduced noise and artifacts.
- The not accessible original image is assumed to have neither distortions nor significant enhancement.

The components of a NR objective quality metric (NROQM) are *preferential* (e.g., sharpness) and *artifactual* (e.g., noisiness) IQ attributes, whose combined assessment is expected to be a reasonable indicator of overall IQ as perceived on average by human subjects. They respectively translate into presence of both desirable (e.g. *detail*) and non-desirable (e.g. *noise*) image features. Among all the relevant image features that can be computed without using the original image, those of easy mathematical formulation are selected based on their perceptual impact on quality and whether they can be accurately detected and quantified. Their estimates can be then translated into visibility of distortions according to psychophysically established relations, and finally integrated using some combination rule to obtain a quality score.

Without a reference with which to compare, NR metrics necessarily rely on some *a priory* knowledge, often in the form of strong assumptions about the feature characteristics, as well as about image formation and distortion processes, thus often following, either implicitly or explicitly, structural approaches in the same philosophy as the SSIM.

---

[46] Yet one can still assume that there exists a high quality "original image", of which the image being evaluated is a distorted representation. It is also reasonable to make a further assumption that such a conjectured "original image" belongs to the set of typical natural images. In fact, the label no-reference really means free choice rather than absence of reference.

Although one can compute more than a hundred features of an image, in most situations ***contrast*** (tone reproduction), ***resolution*** (sharpness or amount of detail), ***noise***, ***clipping distortion*** and ***compression artifacts*** constitute the minimum set required to effectively estimate perceived image quality, with the ability to measure both improvement and degradation [53][56][65]. This represents a significant step towards the development of content-independent, no-reference objective quality models as it sheds light on the key quality factors and how they influence image quality.

3.4.1 Non-desirable or *artefactual* image features

3.4.1.1 Clipping

The term *clipping* refers to a truncation in the number of bits of the image values (luminance and chrominance components) imposed by the arithmetic precision of the process being used, which results in abrupt cutting of peak values at the top and bottom of the dynamic range. Sharpness enhancement techniques can cause clipping, since most of them work by adding positive and negative overshoots to the edges. The simplest clipping measure is a function of the percentage of clipped pixels found in an image.

3.4.1.2 Compression Artifacts

Well known examples that strongly affect the perceived quality of compressed images are *block(ing)* artifacts, which result from a coarse quantization of DCT coefficients –of 8x8 pixel blocks- in JPEG coding, and *ringing* artifacts (a shimmering effect around high contrast edges) in wavelet-coded (e.g. JPEG2000) images [65]. The level of the former can be estimated by the likelihood of detecting artificial horizontal or vertical edges around block borders, while in the latter case it is given by a ratio indicating the deviation of the spectrum of noise filtered out by an *edge-preserving* smoothing filter (e.g. the Bilateral Filter described in Chapter 4) from the white noise spectrum.

3.4.1.3 Random noise

Noise is a random variation in the range domain, which appears in images as a result of random processes linked to capture and transmission techniques. It is most noticeable in smooth regions or regions with smooth transitions, giving the subjective impression that the image is not clean, or that something unintended is superimposed on the image. While in some cases, small amounts of high-frequency noise add to the "naturalness" of textures (in contrast with a plastic or synthetic appearance) and have been found to increase perceived quality, most noise, however, obscures details and reduces quality of the visual information. For images corrupted with AWGN, it has been empirically shown/found that the perceived distortion is proportional to $log(variance_{noise})$. In order to estimate it, most algorithms assume that the reference image contains at least small areas of constant brightness. Hence, whatever observed variation in these areas is nothing but noise. Robust estimators of the variance, such as the *Mean Absolute Deviation* (MAD) have also been used. The problem of noise estimation will be studied in deeper detail in Chapter 4

3.4.2 Desirable or *preferential* image features

3.4.2.1 Sharpness

Sharpness, also referred to as *micro-contrast*, refers to the perceived degree of clarity of detail and contours of an image. A sharp photo of a scene is almost always preferred to a blurry photo of the same scene, which is often the result of poor technique, e.g. camera shake, or poor equipment, e.g. low quality lens. Indeed, it is extremely rare for an entire photo taken by a professional to be blurry. There is always at least some part of the photo that is sharp and in focus. In fact, there seems to be a minimum of feature occurrence necessary just to achieve a "good" visual image. An image should fundamentally be a feature-rich visual representation, ideally associated with a near-perfect sense of clarity.

Under the assumption that edges are (presumably) the most important features in the image source [68], objective sharpness measures typically disregard the dependency on content, spatial resolution, contrast, and noise, and focus on the definition of edges in the spatial domain (e.g., based on local gradient or edge kurtosis), or on the characteristics of the high frequencies in the transformed domain (e.g., based on the maximum frequency of the image I_b, estimated as the number of frequencies whose power is greater than some threshold θ), as a reliable indicator of perceived image sharpness [47].

3.4.2.2 Contrast

Contrast refers to the perceived degree of separation of different tones in an image. Professional photos typically have higher contrast than snapshots. Low contrast photos look washed out. Under the assumption that the response of the HVS depends, not that much on the absolute luminance, but on the relation of its variations to the surrounding background, most objective contrast measures disregard the dependency on factors such as a mental reference image of the object in question, overall luminance, or colour, and define contrast as a measure of relative luminance variation, i.e. $C = \Delta L / \bar{L}$. For example, $\Delta L = L_{\max} - L_{\min}$ or $\Delta L = \sigma_L$ (see Chapter 5). The rationale behind this is that a small difference is negligible if the average luminance is high, while the same small difference matters if the average luminance is low, what is known as *'Weber-Fechner law'* [55]. More complex methods are based on the study of the shape of the intensity histogram.[47] This idea can be extended to more general tone-reproduction properties other than contrast, such as spanning the full range, a bell-shaped distribution, no gaps (no posterization), etc.

For complex images, however, it is difficult to find a consistent definition of contrast. Peli's local bandlimited (i.e., at each spatial position and frequency) contrast [58], is defined as the ratio of the *bandpass* to the *lowpass* filtered version

[47] E.g., a very basic algorithm would include the following steps: 1) compute the luminance histogram; 2) separate the upper and lower parts that contain each a certain percentage of the total energy; 3) calculate the difference between them and normalize by the average luminance.

of the image $c_h(x,y) = a_h(x,y)/l_h(x,y)$, at a given band of spatial frequencies *1/h*. The power (intensity variance) at the lowest frequency represents the average image luminance, classifying it as either a dark or bright image. The level at the highest frequency represents the *micro-contrast*, classifying the image as either blurry or sharp. The levels at middle frequencies measure image contrast at different observation scales.

3.5 Image Quality Improvement

3.5.1 Depiction as Optimization

In order to develop relevant solutions to the problem of image quality improvement, we need first to recognize the complexity of the depiction problem and its optimization dimension: *image quality improvement is essentially an optimization problem that aims at producing the most relevant picture for a given purpose*.[48] For example, for many decades photography has evolved an empirical set of (preferred) reproduction goals, which might include (but of course are not limited to) producing pleasing or preferred pictorial images, reproducing overall appearance of the scene, maintaining contrast relationships between objects, maintaining the original photographers' intent, and predicting visibility of specific objects in a scene.

While this optimization problem should most of the time be solved by the user, the optimization nature of the process requires the design of specific tools for efficient user interaction. There are essentially three strategies to solve this optimization problem: *a)* the user can solve it, *b)* the computer can solve it, or *c)* the solution might involve both user and computer decisions. The general case is mixed: the computer has to take decisions automatically, but the user wants to keep some control and influence the decisions according to the intended use of the image.

Therefore, a central question and the first step in the processing of (pictorial) images (for their improvement) is the intent of the reproduction. But the chosen goal needs not only to be appropriate for the intended use of the image, but they also needs to be realizable with the intended capture and output equipment. For example, the implementation of the above mentioned goals in conventional photographic systems has been driven to some extent by materials considerations, with the relative rigidity of chemical processes preventing them from being tweaked in ways that would have been advantageous with more flexible systems, such as digital systems.

[48] Following [52], we use the term "picture" to describe a visual representation of a visual scene. In contrast to a photograph, such an image is not necessarily optically accurate (e.g. a line drawing). He also notices that most depiction issues are common to realistic and non-photorealistic styles, and that photorealistic rendering is only a special instance of depiction.

3.5.2 Reproduction Goal Choices. Types of realism

Depending on the application, two intents can be clearly differentiated: 1) ***accurate*** reproduction**,** commonly (but wrongly) referred to as *realistic*; and 2) ***pleasant*** reproduction, typically referred to as *preferred*. The latter, based on ***subjective preference***, strives to make the rendered image look as pleasant as possible to the viewer (this is usually desirable in consumer imaging and commercial photography). For the purpose of this present thesis work, from now on we will assume that the image should be accurate and not necessarily pleasant.

From a discussion of the levels of accuracy found in [52], depending on the level of visual coding at which accuracy is defined, accurate reproduction can be profiled to achieve three different goals: a) ***physical (objective) match***, where the image provides the same ***visual stimulation*** as the scene. It is degraded because of the presence of artefactual attributes such as noise or blur; b) ***perceptual (subjective) match***, which strives to make the rendered image as perceptually similar as possible to the original scene (this is usually an implicit goal in consumer imaging and image synthesis applications); and c) ***functional (cognitive) match***, which seeks to preserve or enhance the information of an image, usually details at all regions and all luminance levels, as is most often requested in medical imaging, satellite imaging, and archiving.

Considering aesthetical match as a (subjective) level of accuracy, we can say that ***physical***, ***perceptual***, ***aesthetic*** and ***functional*** approaches to image quality respectively result in ***identical***, ***natural***, ***pleasant*** and ***useful*** images.[49]

Table 3.1 Accuracy levels and related objectives

a. ***Physical / objective*** match.
   The image produces the same ***visual stimulation*** as the scene
   i. *Spectral*
   ii. *Exact (absolute)*
   iii. *Linear (relative)*
   iv. *Colorimetric (photographic realism?)*

b. ***Perceptual / subjective*** match.
   The image produces the same ***visual response*** (appearance) as the scene.
   i. *Exact (brightness-colourfulness)*
   ii. *Relative (lightness-chroma)*

c. ***Functional / cognitive***.
   The image preserves or improves the ***visual information*** in the scene.
   i. *Detailed (enhanced details)*
   ii. *Abstracted*

[49] One may regard aesthetics as a kind of functional match with a high subjective component.

3.5.2.1 Physical accuracy (objective match)

The image provides the same visual stimulation as the scene. This means that it has to be an accurate point-by-point representation of some physical measurement, such as the spectral irradiance, at a particular viewpoint in the scene.[50]. Nevertheless, with the state-of-the-art in digital image capturing, it is feasible to sample spectral irradiance with high precision and resolution. Moreover, with accurate image synthesis techniques, digital images can be accurate numerical simulations of light reflection and transport. However, and in spite of its great utility for quantitative analysis in a wide range of design and engineering applications, physically accuracy for observable realistic images of natural scenes is rarely feasible, practical (n)or even appropriate for a number of reasons.

First, such images are rarely realizable one existing media. For example, conventional displays cannot, in general, reproduce the original spectral irradiances. Consider a photograph of an outdoor scene in bright sunlight, which is then printed and viewed in an indoor environment. It is physically impossible to achieve an absolute match so that the measured energy coming off the print is the same as the original outdoor scene. Second, it is overkill if one's job is to create images for human observers, since their visual limitations are not taken into account. For example, colour imaging technology takes advantage of the trichromatic nature of vision to reduce the requirements for describing colours from their full spectral representations to their metameric RGB or CMYK equivalents, as described in [55]. And third, because the differences between the scene and the reproduction viewing conditions cause very different states of adaptation of the HVS, physical accuracy alone does not guarantee that the resulting image will have an appropriate appearance when displayed. For example, consider the reverse situation, where the original photograph is now taken indoors and then reproduced and viewed outdoors. In such a case the print would have to be unreasonably dark to match the absolute attributes of the indoor scene.[51]

For most general imaging applications, a relative match is preferred so that the relationship between objects in the scene is held constant. Second, the reproduction should be not only *physically correct* but, what is more important, also *perceptually equivalent* to the represented scene. By incorporating the observer's visual system in the reproduction process, it is possible to take advantage of the limitations of vision to simplify the task and knowledge of what is relevant to produce more compelling reproductions.

---

[50] Depending on whether absolute or relative measurement values are taken, the image is respectively said to be an *exact* or a *linear* reproduction of the scene.

[51] Briefly, images rarely render a precise replica of the original scene because of their many inherent limitations compared to the real optical flow: they are flat and of limited extent, field of view, gamut and contrast. Besides the limitations of the medium, the reproduction is typically viewed at a different (usually smaller) size, from a different perspective and, more important, in a different context than the original.

3.5.2.2 Perceptual accuracy (subjective match)

For many applications, images should not only be physically correct but also perceptually equivalent to the represented scene. That is, the image and the scene it represents should provoke the same visual response (i.e. an appearance match) when each is viewed under specific conditions. This task is facilitated by colour appearance models [55].[52]

Essentially, given an input image and viewing conditions, an image appearance model can provide perceptual attributes of each pixel and describe human perception of the image. The inverse model can take the output viewing conditions into account and thus generate the desired output perceptual effect. For example, by developing models of how the HVS adapts to the vast ranges of light energy found in different scenes, researchers have been able to design (tone mapping) algorithms that reproduce the appearance of high dynamic range (HDR) scenes within the limitations of low dynamic range (LDR) display devices (see [61] for a recent review).

While this approach allows adopting a perceptual image processing for application purposes involving image quality assessment (both, FR and NR) and image rendering [53], it has also several drawbacks. On one hand, complete forms of perceptually-based algorithms are necessarily extremely complex in order to account for the observed phenomena and are still too low for interactive applications; they require data about scene's and reproduction's viewing conditions, which is not available from consumer cameras; the visual models on which they are based do not explicitly account for the many cognitive mechanisms impacting image appearance, such as memory colour, perceptual constancy discounting the illuminant and object recognition (these may be implicitly included, though, as they are present in experiments). On the other hand, it is unclear that photo-realism is necessary or even desirable in a wide range of graphics applications, since adopting it as a standard for visual realism in computer graphics, classifies most renderings as failures, yet says nothing about their obvious utility in many domains. Often, non-photorealistic pictures can be more effective at conveying information, more expressive or more beautiful.

---

[52] Notice that CIE colorimetry is only strictly applicable to situations in which the original and reproduction are viewed in identical conditions. By their very nature, the images produced or captured by various digital systems are examined in widely disparate viewing conditions, from the original captured scene, to a computer display in a dim room, to printed media under a variety of light sources. When this is the case, it becomes necessary to specify the actual color appearance. Complete specification requires five perceptual dimensions: *brightness*, *lightness*, *colorfulness*, *chroma* and *hue*. For most imaging applications, it is often desirable to attempt for a *lightness-chroma* match rather than the absolute *brightness-colorfulness* match.

[53] With proper calibration, the images can be predictive visual simulations that accurately show what an observer would see if they were in the scene and, if the visual models can be validated, then the images could be used for quantitative visual analysis in a wide range of design and engineering approaches.

3.5.2.3 Functional accuracy (cognitive match)

Neither at the physical nor even at the perceptual, but at the cognitive level, for most applications the final goal of an image is to provide the same visual information as the scene it represents. This naturally yields to a functional, objective definition of image quality in terms of how well the desired information about the scene can be extracted from the image, measured by the performance of some "observer" on some specific task. [54]

This is the approach pursued by Janssen in [53]. Formally, he regards 1) images as carriers of visual information about the outside world; 2) images as input to visual perception, which together with cognition and action, constitute human interaction with the environment; and 3) image quality as the adequacy of the image as input to the vision stage of the interaction process. For these stages to be completed successfully, the image should in general satisfy two main requirements: (1) the internal representation of the image should be sufficiently precise; and (2) the match between the internal representation of the image and "knowledge of reality" as stored in memory should be close. Janssen refers to the degree to which an image satisfies these two requirements as the ***usefulness*** (that is, the precision of the internal representation of the image) and the ***naturalness*** (that is, the degree of match between the internal representation of the image and representations stored in memory) of the image, respectively.

According to this, the quality of an image is then defined to be the degree to which the image is both useful and natural. Observe that the sets of requirements that one needs to impose upon an image in order to maximise the usefulness or the naturalness of this image will in general not coincide. For example, detection or discrimination of objects in an image may require "exaggeration" of certain features of this image, resulting in a less natural reproduction of the image.

The beauty of such a functional definition of image quality is that it admits a wide range of rendering styles from physically-based simulation through photo-realism, to more abstract approaches such as non-photorealistic rendering. Moreover, the concept of quality here presented is formulated independently of modality, which opens the possibility to apply it to, for example, sound or speech quality.

---

[54] Here information means knowledge about the meaningful properties of objects in scene, such as their shapes, sizes, positions, motions and materials that allows an observer to make reliable visual judgements and to perform useful visual tasks. For example, a good illustration/drawing may actually be better at conveying information to an observer than a physically accurate or photo-realistic image, as it can eliminate irrelevant details, facilitate visual segmentation and grouping, show viewpoints that would be difficult (or even impossible) in a photography or even make use of "special effects" like artificial transparency to depict important features that would be hidden in photographs. One example of functional realism in computer graphics are the images used in flight simulators. The proof of the realism of these images is that they allow the observer to learn skills that then transfer into the real world [52].

3.5.3 Unified framework for accurate reproduction

Once we recognize that the acquired image data are used to reproduce a generalized view of the original, the connection between image reproduction and image interpretation becomes clear: the image reproduction process will be more successful if we interpret the original and apply appropriate transformations. For example, interpreting the shape of an original object can improve subsequent renderings from different perspectives; illumination estimation can improve colour rendering, measuring motion can remove motion blur. In general, imaging systems that can interpret the original image data will have better image reproduction capabilities.[55]

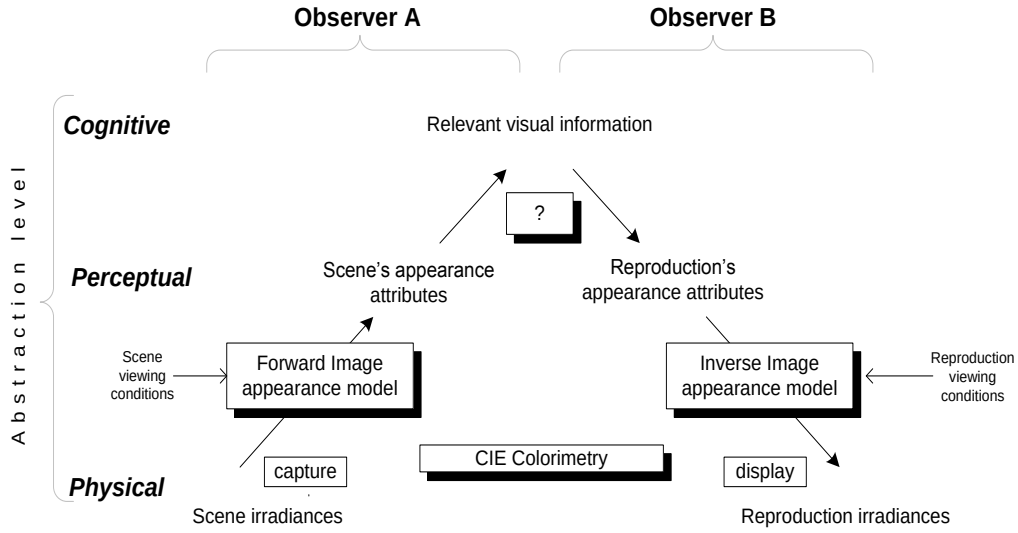


Figure 3.11. Flow chart for the proposed computational model. The model describes image quality in terms of physical processes in the imaging system and psychophysical processes in the hypothetical scene and display observers that affect the fidelity of the displayed image to the scene. The model has two main parts: the *Capture* and the *Display* models. The former, which corresponds to the scene observer *A*, processes an input image to encode the perceived features. The latter, which corresponds to the reproduction observer *B*, then takes this encoded information and reconstructs an output image. The model must be inverted in order to produce equivalent appearances under the viewing conditions of the display device. This procedure does not "undo" the former processes, since the visual models differ for the original scene and the display. The reconstruction process creates instead an output image that reproduces either the *physical*, *perceptual* or *cognitive* content of the input image according to a reproduction intend and subject to the limits possible on a given display device, so as to maximize the relevant mutual visual information shared by both observers. Posing the problem this way, we reinforce the analogy between *capture* and *perception*, and between *rendering* and *depiction*. Capture/perception is an *analysis* process in that the relevant *intrinsic* characteristics such as noise, object surface reflectance and shading, are obtained from the observed image. Rendering/depiction is a *synthesis* processes, in that a new image is constructed from its description, in which some intrinsic components are enhanced (e.g., surface reflectance) and other are discarded (e.g., noise).

[55] The emphasis on the importance of image interpretation is an extension of current practice; modern electronic imaging systems already include control systems that make certain inferences about the scene. Exposure value systems analyze image intensity, white balance systems analyze the color distribution in the image, and focus systems measure the distance to a principal object. We believe that electronic imaging systems of the future will derive and encode much more information about the physical characteristics of the scene.

3.5.4 Analysis performed in different communities

Artists and other picture makers have developed a rich set of techniques to produce effective pictures. We believe that computer graphics may learn a lot from this large body of knowledge, as well as from the analysis performed in the perception community. The task is not easy because the craft is often elusive or expressed in terms that are not easily translatable to algorithms [3].

3.5.4.1 Analysis done in the Artistic community

As remarked before, pictures have limitations compared to the real world: they are flat, of finite extent, have a limited field of view, represent the scene from a single point of view, are often static, and they have limited gamut and contrast, just to mention a few. But they can also be *compensated* for, using pictorial techniques to convey the missing cue or dimension using a different mode (e.g., flatness can be compensated for by accentuating the contrast at the occluding silhouette).[56] Pictorial techniques, such as photographic lighting, processing, or *dodging and burning*, remain largely unexplored, although they are fundamental parts of effective depiction, and can prove a key aspect of the digital photography revolution. They not only can alter the picture to address the limitations of the medium, but may also aim at more effective pictures: they omit irrelevant information, and emphasize the important one, often distorting it in the service of communication. When well designed, such techniques capitalize on humans' facility for processing visual information and thereby improve comprehension, memory, inference, and decision making.

We are mostly interested in 2D→2D-attribute compensation techniques, which include standard image controls such as dodging and burning [44], contrast/brightness or colour modification, e.g. [61]. In Chapter 5 we consider in more detail *tone mapping*, also a 2D→2D attribute technique, which specifically copes with the limited gamut and contrast of common reproduction media . Specifically, we present techniques for coping with different absolute intensities and contrast management. They require the development of appropriate operators and interactive image editing techniques based on perceptual phenomena. To fully account for the diversity of picture styles and to understand the mental processes involved, one has to think of depiction as the inverse of the inverse problem. Indeed, representing a given scene consists in producing a picture that induces a similar impression to beholders as they would have in front of the real scene [3].

[56] It should be mentioned that limitations can also bring important richness to pictures, and can therefore be accentuated (e.g., black-and-white photography is often considered more artistic despite its missing color dimension). They can also be eliminated, usually through technological -rather than pictorial- solutions that extend the pictorial medium and reintroduce the missing dimension of the visual experience (e.g., the use of stereo-pairs eliminates some of the limitations related to flatness).

3.5.4.2 Analysis performed in the perception community

It often occurs that an interesting 3-dimensional (3-D) scene looks perfectly fine to the visual perception of a photographer, but appears very dim or weakly contrasted when captured in an analog or digital camera. A camera always faithfully snapshots the physical luminance of a scene (weather linearly or nonlinearly). It is the human visual perception that has regulated and better conditioned the actually dim scene, and maintained stable perceptual performance. For example, we see a red apple as red under illuminant with very different colour temperatures, although the physical stimuli have very different objective chromaticities. This ability to discount the accidental conditions (e.g., the colour of the illuminant) and to extract (scene's) invariants (e.g., object reflectance), known as ***perceptual constancy***, is crucial in studying vision and the complex dualism of pictures.

When looking at a picture, constancy might not operate the same way as when looking at the scene. For example, chromatic adaptation does not function equally. This is why white balance is needed for video cameras, or why different films are required for outdoor photography and for indoor photography without flash. Indeed, when we look at a picture, our visual system adapts to the colour of the illuminant of the room in which we look at the picture. In contrast, we are able to discount the intensity of the illuminant in a picture, as demonstrated by Figure 3.12.

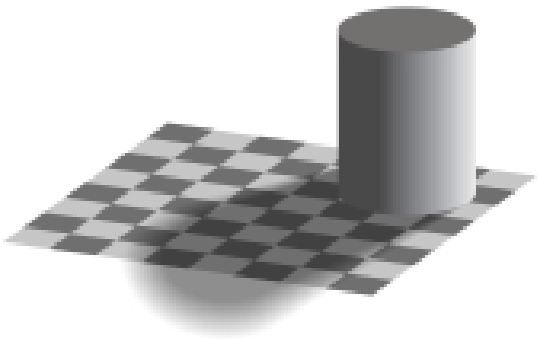


Figure 3.12 In this picture, the white cells in the shadow of the cylinder have the same grey level as the black cells in full light. After an illusion by Ted Adelson.

While invariants are often represented directly (not only because invariants are easier for us to consciously access, but also because invariants are by nature a "better," or at least more immutable representation), most pictures are hybrid, and managing the balance between extrinsic (*"what I see")* and intrinsic (*"what I know"*) properties is one of the keys to good depiction.[57] A common way to solve the dilemma between extrinsic and intrinsic characteristic is to choose the depiction such that the extrinsic characteristics match the intrinsic ones. For example, in cinema and photography, by using a fill light that illuminates the shadowed areas. Note that this means choosing the depiction situation (additional light source) in order to improve the picture: the 2D picture influences the depicted scene.

[57] It suffices to read the opposite statements made by the 19th century painter Turner who claimed, *"My business is to paint not what I know, but what I see"* and by the 20th century Picasso who declared, *"I do not paint what I see, I paint what I know"*.

3.5.4.3 Analysis performed in the information-theoretic community

Image quality can be defined objectively in terms of the performance of some "observer" (either a human or a mathematical model) for some task of practical interest. For example, for scientific and medical purposes, it can be defined in terms of how well desired information can be extracted from the image.[58] Following the ideas presented in Section 3.3.2, here we propose to extend the information fidelity paradigm from the FR to a NR approach. Thus, instead of regarding the reference and test images respectively as the input and output of a communications channel, here we consider the channel itself [62]:

> *Perceived image quality is proportional to Shannon information capacity, which is a function of both image sharpness and noise.*

Because of the gaussianity assumption used to model both image and noise, equation (3.12) corresponds to the channel capacity as given by Shannon's classic equation for the information transmission capacity C of a data channel: $C=W\log_2(S/N+1)$, where W is the channel spatial bandwidth, which corresponds to image sharpness [59], and S/N is the signal-to-noise ratio (SNR), which measures the maximum number of distinguishable levels. These can be either measured from the capture device (in which case, we are measuring the perceived quality of the 'system'), from a reference image (FR) or even estimated from the image itself (NR). In practice, it is better to express C in bits/steradian

$$W = \frac{P}{\Omega} = \frac{A_{im} / p^2}{A_{im} / r^2} = \frac{r^2}{p^2} \qquad\qquad C = \frac{r^2}{p^2}\log_2\left(\frac{S}{N}+1\right)$$

where W is now the number of *effective pixels* per steradian, A_{im} is area of the image, p the *effective pixel dimension* [60], and r is the viewing distance.

This formulation allows for explicitly decoupling frequency distortions and noise injection. This decoupling, implicit in the structural similarity approach, enables both effects to be quantified and the performance of restoration algorithms to be assessed (these usually introduce frequency distortions when attempting to noise reduction). In fact, this theoretical approach to image quality assessment further leads us to consider image quality improvement in terms of: increasing image sharpness (W), improving tone reproduction (S) and reducing the noise level (N), in order to increase the amount of perceived detail. In fact, these are the three separable dimensions of image quality already mentioned at the beginning of this chapter.

---

[58] In general, the tasks can be divided generically into *classification* and *estimation* tasks. In medical applications, an example of a classification task would be lesion detection, while an estimation task might be determination of the blood pressure.

[59] The Modulation Transfer Function (MTF) describes the ability of an imaging system to capture image contrast over a range of spatial frequencies. These curves are summarized by a single value, the spatial frequency at which the amplitude falls to 50% of the highest amplitude (MTF50). Sharpness is typically well correlated with this value.

[60] E.g., estimated as the std. deviation of the point spread function resulting from the combined effect of the lens, display and eye blur.

It is important to notice that, even under ideal capturing, transmission and reproduction processes, the amount of information carried by the light forming the image is necessarily limited by two fundamental properties of light: the ***SNR*** is limited by ***photon noise,*** due to the discrete stochastic nature of photon production and counting ([59], Chapter 4, Section 4.9.1 *Sources of Noise*); and the resolution is limited by ***diffraction-limited blurring***, due to the ondulatory nature of light [61]). This wave-particle duality, which relates to the uncertainty principle[62], introduces a spatial vs. tonal resolution trade-off when the size of an imaging sensor array (and thus the total amount of light falling on it) is fixed.[63]

Notice also that the above formula measures image quality in terms of image detail, which is the photographic equivalent to information capacity. However, in order to obtain a perceptually meaningful measure, the limitations of the HVS should be incorporated: its power is limited in both the spatial domain (as measured by the CSF) and tonal or range domain. The HVS is itself a source of noise, which prevents it to distinguish more than about 100 levels in a reflective image [68]. According to this, a better definition of the capacity would be as follows

$$C = \sum_h C_h\,, \qquad C_h = \sum_{x,y} \log_2\left(\frac{c_h(x,y)}{N + N_{HVS}} + 1\right) \text{ (bits)}$$

where $\boldsymbol{C_h}$ is the channel capacity *per band*, $\boldsymbol{c_h(x,y)}$ is Peli's local band-limited contrast [58], and $\boldsymbol{N_{HVS}}$ is the equivalent noise of the HVS.

Shannon capacity may well become accepted as a metric for measuring image quality when (1) devilish details in measuring W, N, and S are worked out, (2) the concepts become more familiar, and (3) perceptual testing (relating C to perceived image quality) is performed.

The approach contributes to a more meaningful description of image quality, it includes band limitation and noise, and opens a way for a more standardized evaluation. Not only the primary process of imaging, but also methods for image processing, which affect the composition of spatial frequencies or filter the noise, appear to be comparative under such an approach.

---

[61] The on-axis blurring of a circular diffraction-limited imaging lens is characterized by a point-spread function in the form of an Airy pattern [63].

[62] The uncertainty principle states that that the values of certain pairs of conjugate variables (position and momentum, for instance) cannot both be known with arbitrary precision. That is, the more precisely one variable is known, the less precisely the other is known.

[63] First, consider photon noise. As pixel size shrinks, the mean photon count at the photodetector falls. The Poisson variance of the photon signal at the pixel equals the mean photon count, so that the signal-to-noise ratio (SNR) of the photon signal decreases. Thus, reducing pixel size inevitably increases image noise. Second, consider diffraction: As pixel size shrinks one can increase spatial sampling density (pixel pitch). But image resolution does not improve without bounds because the spatial detail in the image is limited by diffraction, i.e., the spatial spread caused by light passing through a finite aperture. There has been relatively little analysis of the tradeoffs in image quality as one chooses between spatial resolution and sensitivity improvements [46].

About white Gaussian noise assumption

Besides the mathematical convenience for using such a model, white Gaussian noise is assumed because it is the worst type of noise one can have, insofar as perturbing the message is concerned, because it is the one with greatest entropy for a given noise power.[64]

3.5.5 Improvement as normalization

Image processing towards achieving a given objective target may be regarded as ***image normalization***. Good examples are histogram equalization, white patch and gray world colour correction, etc.[65] Such a transform is further constrained by other objectives, such as naturalness or detail preservation. The final goal may be achieved completely (e.g. complete chromatic adaptation) or only partially (e.g. incomplete chromatic adaptation).

In general, visually optimized images seem to be more tightly clustered about a single mean value $\bar{I}$ and have much higher standard deviations $\bar{\sigma}_f$, as shown in Figure 3.13. These results support the idea that visual optimization centers the data mean on the mid-point of the image dynamic range and spreads the signal excursions out across the dynamic range to a maximal extent while at the same time limiting any over- and under-shoots spatially. This overall trend relates to most efficiently occupying the data space with the actual image data. In general, visually optimized images are improved in terms of both regional lightness and contrast with the latter being the most strongly affected [54].

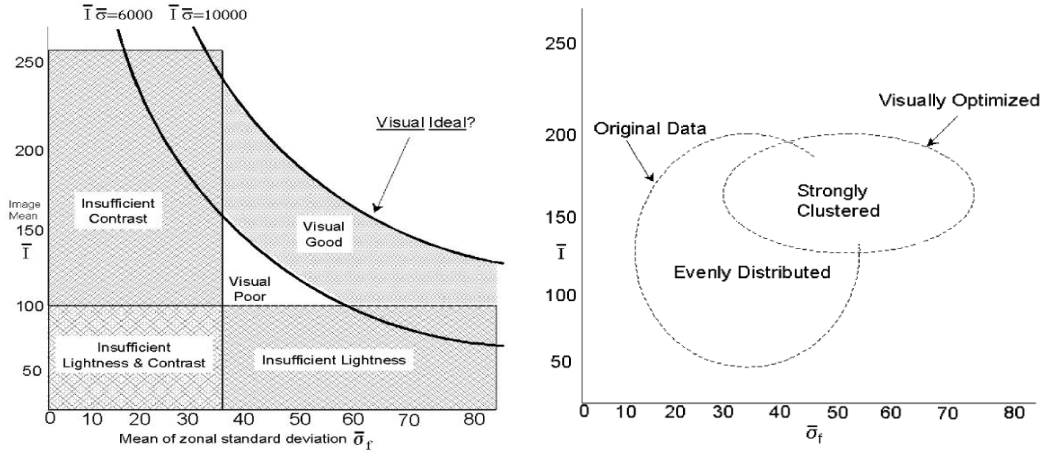


Figure 3.13. Second order statistical characterization of images based on visual appearance. a) initial hypothesis (contrast and lightness); b) actual overall optimization trends. Adapted from [54].

[64] Given an arbitrary type of noise (e.g. impulsive noise, or white noise that has undergone a nonlinear process), we can calculate the power of a white Gaussian noise having the same entropy as the given noise. This power, namely *N=1/(2*pi*e) exp(2*H)*, where *H* is the entropy of the given noise, will be called the *entropy power* of the noise. A noise of entropy power *N* acts very much like a white noise of power *N*, insofar as perturbing the message is concerned.

[65] Although it seems unlikely that the optimal output histogram is completely independent of image content, the principle of specifying target output image tone characteristics has been incorporated into recent tone-mapping algorithms intended to improve upon histogram equalization. In these algorithms, the output histogram varies with an analysis of image content.

3.6 Proposed approach

We build on the theoretical approach to image quality assessment from the information-theoretic community to further consider image quality improvement in terms of:

1. increasing detail by reversing the blurring process (H),
2. improving tone reproduction by discounting the illuminant (I), and
3. reducing the noise level (N),

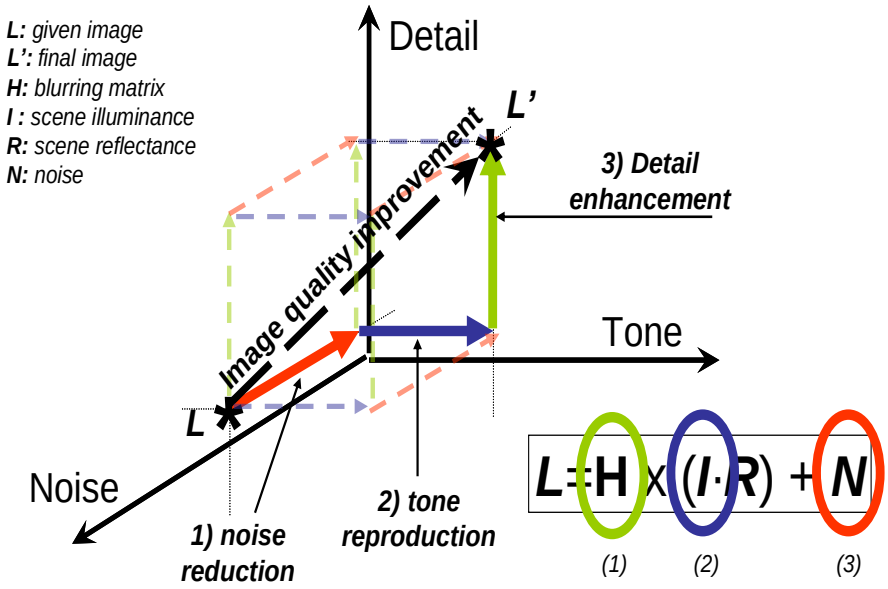


Figure 3.14. Proposed IQ improvement approach. It compromises 1) increasing image *detail*; 2) improving *tone reproduction*; and 3) reducing the *noise level*.

In fact:

- these are the three separable (orthogonal) dimensions of image quality already mentioned at the beginning of this chapter.
- this scheme can be tuned to achieve physical, perceptual or cognitive accurate reproductions.

In next chapter, we present edge-preserving smoothers in the context of image noise removal. As we will see, these are a very simple, yet powerful tool to separate an image in its intrinsic components.

3.7 Summary

Image quality assessment is becoming an increasing research area, with a growing number of emerging approaches from many fields. This chapter has illustrated how close, yet how far we might actually be from achieving automatic image quality assessment and improvement. The ideas presented here spring from reevaluation of our knowledge about image and distortion structure, high-quality images, the human visual system, and the reproduction intent. Together, they provide a unifying framework in which to develop techniques for image quality improvement.

The material was selected based on their applicability to image quality improvement, with a threefold purpose: first, to introduce the fundamentals concepts and to explain the most relevant engineering problems. Second, to provide a proper framework by means of a broad but unifying overview of state-of-the-art leading algorithms that approach the problem from different assumptions, with focus on those applicable to image quality improvement. By adding to our understanding of what is to be measured when dealing with images and by strengthening the bridge between the objective (physical) and the subjective (visual) aspects of many image processing issues, these ideas have clarified the meaning of image quality and thus have enhanced our ability to obtain it. Third, to provide new directions of future research, by presenting new emerging paradigms that are still conceptually new enough that may be further improved.

Image quality based on aesthetical preferences is out of the scope. This chapter takes an approach based on accuracy, where we distinguish three levels: *physical* (objective) match, *perceptual* (subjective) match and *functional* (cognitive match). These respectively result in *identical*, *photorealistic* and *detailed* images, providing a unifying framework.

In general, three types of knowledge can be used in the design of image quality assessment and improvement methods: knowledge about the human visual system (HVS); knowledge about high-quality images; and knowledge about image distortions:

a) **Knowledge about the HVS** can be further divided into *bottom-up knowledge* and *top-down assumptions*. The former includes computational models that have been developed to account for a large variety of physiological and psychophysical visual experiments. The latter refers to those general hypotheses about the overall functionalities of the HVS. For example, the structural similarity principle introduced in Section 3.3.1 assumes that the HVS is adapted to separate structural information from nonstructural information from the visual scene. The information theoretic approach presented in Section 3.3.2 is another example, where the HVS is considered as an information communication channel and mutual information is

employed as a measure for information fidelity. In the case of application-specific image quality assessment, it is also sensible to make top-down assumptions about visual tasks, such as object detection.

b) **Knowledge about high-quality images** can be either *deterministic* or *statistical*. In the case of FR image quality assessment, which was discussed in Sections 3.2 and 3.3, there is a single high-quality original image that is completely known in a deterministic fashion. In the case of RR quality assessment, the knowledge is statistical, in the form of a set of selected statistical features, but still about a single high-quality original image. In no-reference (NR) quality assessment, however, the assumed statistical knowledge describing high-quality image is not restricted to a single original image, but rather, expressed the probability distribution of all high-quality natural images that fall within the space of possible images. For a given test image, the quality assessment work is carried out by measuring its departure from such a probability distribution of natural images. In this situation, image quality degradation is equated with "*unnaturalness*", which is, no doubt, a top-down assumption about how the HVS looks at the world. Indeed, this outlook may be justified from the viewpoint of computational neuroscience. In that context, the "efficient coding" principle states that *the role of early biological sensory systems is to remove redundancies in the sensory input, resulting in a set of neural responses that are statistically independent* [40]. According to this, modeling the HVS and modeling natural image statistics can be considered as dual problems, since the former must be highly adapted to the latter.

c) **Knowledge about image distortions** is also a useful source of information for the design of image quality measures, especially in the case of application-specific image quality assessment where efficient algorithms may be developed by directly evaluating the strength of a few specific types of image distortions. Examples are given in Section 3.4. The case of general-purpose image quality assessment, however, is much more complicated, since the specific types of distortions are not known beforehand and universal distortion models are not available yet.

Based on presented ideas, we envision a three-step process of improving an image quality: (1) detect the presence of unwanted artefactual attributes (such as noise or blur) (2) accurately estimate a quantitative description of relevant parameters (such as noise variance or amount of blur) (3) eliminate them and recover the underlying original image in a natural way (i.e., select the most appropriate algorithms for the particular degradation found, on the basis that it will result in a "natural image"). Points (1) and (2) relate to quality *assessment*, while point (3) relates to quality *enhancement*.

With respect to the latter, it is noticed that recorded colour images differ from direct human viewing by the lack of dynamic range compression and colour

constancy. In chapter 5, research is summarized which develops the center/surround Retinex concept originated by Edwin Land to achieve dynamic range compression, colour constancy, and colour rendition and, eventually produce a perceptual match to direct observation of the scene.

Extensions and Future Work

The limited space of this chapter has allowed us to introduce the basic problems, ideas and exemplar approaches to image quality assessment and improvement. The wide range of applications extends the field of image quality assessment into other dimensions including, but not limited to, image compression, communication, acquisitions, printing, display, restoration, enhancement, denoising, segmentation, detection, and classification of photographic, medical, geographic, satellite, and astronomical images. The general methods discussed in this chapter are certainly extendable to these areas.

Improving the performance of described methods is also possible if they are to be applied to specific applications. First, the distortion types are usually constrained and predictable for given application environments, and the measures that can directly quantify these application-specific distortions may provide useful indications of image quality. Second, specific applications are typically associated with specific visual tasks. For example, the ability to visually detect certain objects would be a very important factor for assessing the quality of medical images.

Perceptual image quality is not a stand-alone research topic. In fact, we view it as the core of a much broader field: perceptual image processing. It is desirable to incorporate perceptual image quality with the several types of image processing applications to build perceptually optimized image processing systems. As mentioned in Section 3.2.1, the MSE is still used everywhere, not only to evaluate but also to optimize a large variety of image-processing algorithms, due to its mathematical convenience. A worth direction of research work is to replace it with perceptual meaningful measures. Moreover, image processing may greatly benefit from perceptual approaches. While there has already been some related work [67], it is still in very preliminary stages, and there is, no doubt, a great deal of room for improvement to be explored in the future.

Finally, despite of the high interest in scientific and medical purposes, image quality in terms of visual information capacity as given by Shannon's formula has almost not been studied before. The idea is here presented somewhat informally. Future work would include a more rigorous theoretical formulation.

REFERENCES

[44] ADAMS, A. *The Print. The Ansel Adams Photography series*. Little, Brown and Company. 1983.

[45] BARTLESON. J The combined influence of sharpness and graininess on the quality of colour prints. *Journal of Photographic Science*, Nr. 30. 1982.

[46] CATRYSSE, Peter B.; WANDELL, Brian A. Roadmap for CMOS image sensors: Moore meets Planck and Sommerfeld. En *Digital Photography*. 2005. p. 1-13.

[47] CAVIEDES, Jorge E.; OBERTI, Franco. No-reference quality metric for degraded and enhanced video. *Visual Communications and Image Processing 2003*. International Society for Optics and Photonics, 2003. p. 621-632.

[48] DAMERA-VENKATA, Niranjan, et al. Image quality assessment based on a degradation model. *IEEE transactions on image processing*, 2000, vol. 9, no 4, p. 636-650.

[49] ENGELDRUM, P.G. Image Quality Modeling: Where Are We? *IS&T's PICS Conference,* 1999.

[50] FAIRCHILD, Mark D.; JOHNSON, Garrett M. iCAM framework for image appearance, differences, and quality. *Journal of Electronic Imaging*, 2004, vol. 13, no 1, p. 126-138.

[51] FERNANDEZ, S.R., FAIRCHILD, M.D. and Braun, K. Analysis of Observer and Cultural Variability while Generating "Preferred" Colour Reproductions of Pictorial Images. *Journal of Image Science and Technology*, Vol. 49, No. 1, January/February 2005.

[52] FERWERDA, James A. Three varieties of realism in computer graphics. *Human Vision and Electronic Imaging*. 2003. p. 290-297.

[53] JANNSEN, R. *Computational image Quality*. SPIE Press, 2001.

[54] JOBSON, Daniel J.; RAHMAN, Zia-ur; WOODELL, Glenn A. The statistics of visual representation. 2002.

[55] JOHNSON, G.M. and FAIRCHILD, M.D. Visual psychophysics and colour appearance. in Sharma, G. ed. *Digital Colour Imaging Handbook*, CRC Press LLC, 2003, chapter two.

[56] KEELAN. B.W., Handbook of Image Quality: Characterization and Prediction, Marcel Dekker, New York, NY 2002.

[57] LI, X. Blind Image Quality Assessment. Digital Video Department, Sharp Labs of America.

[58] PELI, Eli. Contrast in complex images. *Journal of the Optical Society of America, A, Optics, Image & Science*, 1990.

[59] PLANCK, M., *Über das Gesetz der Energieverteilung in Normalspektrum.* Ann. d. Phys., 1901. 4: p. 553.

[60] POYTON, C. Frequently Asked Questions about Colour. Available at http://www.poynton.com/PDFs/ColourFAQ.pdf

[61] REINHARD, Erik, et al. Photographic tone reproduction for digital images. *ACM transactions on graphics (TOG)*, 2002, vol. 21, no 3, p. 267-276.

[62] SHAW, R. The Application of Fourier Techniques and Information Theory to the Assessment of Photographic Image Quality. *Photographic Science and Engineering*, Vol. 6, No. 5, Sept.-Oct. 1962, pp.281-286.

[63] SOMMERFELD, A., *Mathematische theorie der diffraction.* Math. Ann., 1896. 47: p. 317-374.

[64] SHEIKH, H.R, BOVIK, A.C. and de Veciana, G. An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics. *IEEE Transactions on Image Processing*, Vol. 12, No. 12, December 2005.

[65] WANG, Zhou; BOVIK, Alan C. *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

[66] WANG, Z., BOVIK, A.C. and SHEIKH, H.R. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions On Image Processing*, Vol. 13, No. 4, March 2004.

[67] WANG, Z., BOVIK, A.C. and SHEIKH, H.R. Structural approaches to Image Quality Assessment. In Bovik, A.C. ed. *Handbook of Image and Video Processing*. 2nd ed. Elsevier Academic Press, 2005, chapter 8.3.

[68] WINKLER, S. *Vision Models and Quality Metrics for Image Processing Applications*. 2000.

[69] WU, H.R. and RAO, K.R. *Digital Video Image Quality and Perceptual Coding*. CRC Press, Taylor & Francis Group, 2006.

[70] ZHANG, Xuemei; WANDELL, Brian A. A spatial extension of CIELAB for digital color-image reproduction. *Journal of the Society for Information Display*, 1997, vol. 5, no 1, p. 61-63.

[71] http://mathworld.wolfram.com/

[72] http://white.stanford.edu/~brian/scielab/introduction.html

[73] Video Quality Experts Group. http://www.vqeg.org

[74] The Design of High-Level Features for Photo Quality Assessment

Chapter 4

EDGE-PRESERVING IMAGE SMOOTHING

Edge-preserving image smoothing has recently emerged as a valuable digital darkroom tool for the task of simplification of visual information, with a variety of applications in computer graphics and image processing. The challenge is to preserve important features, such as homogeneous regions, discontinuities, edges and textures, as much as possible. Its application for image denosing has been extensively studied for decades in the fields of computer vision, image processing and statistical signal processing because of obvious practical importance (whenever it is required effective noise suppression to produce reliable results) as well as its theoretical interest: being perhaps the simplest of inverse problems that image processing researches have studied over a long time, it provides a convenient platform to examine natural image models and signal separation algorithms over which image processing ideas and techniques can be assessed. The variety of reference sources quoted in this chapter is evidence for this fact. Indeed, numerous contributions in the past 50 years or so addressed this problem from many and diverse points of view. Statistical estimators of all sorts, spatial adaptive filters, stochastic analysis, partial differential equations, transform-domain methods, splines and other approximation theory methods, morphological analysis, order statistics, and more, are some of the many

directions explored in studying this problem. In this chapter, we have no intention to provide a survey of this vast activity. Instead, we concentrate on one specific approach towards the image denoising problem that we find to be highly effective and promising: even though they may be very different in tools it must be emphasized that a wide class, regardless of implementation, share the share the same basic idea: denoising is achieved by averaging.

Restoring a signal is typically solved via either the Bayesian approach (already presented in Ch. 2) or filtering, though there are other approaches. Here we focus on the latter because it requires minimal assumptions. In the broadest sense of the term "filtering", perhaps the most fundamental operation of image processing, the value of the filtered image at a given location is a function of the values of the input image in a small neighbourhood of the same location. Filters are roughly grouped into linear and nonlinear types. It is well known that the classical approach to denoising via linear filters (i.e., convolving with constant coefficient windows) cannot handle such a problem since both noise and the mentioned image features contain high frequencies. *Edge-preserving* denoising, requires then adopting a higher order characterization, more localized transforms (wavelets) and/or varying weights depending upon local image structure, resulting in nonlinear algorithms, more effective than linear ones, but also more difficult to analyze, formulate and predict.

While edge-preserving regularization has provided an abundant literature in the last two decades, it is still not easy to see the advantages of the various approaches, and the relations between different methods are only partly understood. In fact, there are only few strategies that combine different approaches and allow further generalizations. However, the integration of several approaches that rely on different mathematical tools (e.g. functional minimization, nonlinear PDEs, statistics and data analysis) is essential for obtaining high-quality results in real-life applications. With this background, the denoising community does perhaps not need more methods, but rather a common framework within which the existing methods can be described and new ideas on improvements are also more likely to appear.

This chapter contributes in this direction by studying several methods and their relations, in order to end up with a better understanding of each of them. We focus on the relationship between regularisation techniques, nonlinear diffusion filtering [81], adaptive smoothing [101], mode filtering, mean shift [89], kernel regression [115], M-estimators from robust statistics [92], bilateral filter [116] and non-local means filtering [83]. Although these methods seem very different at the first glance and originate in different mathematical theories, they are in fact intrinsically connected. On the one hand, each method excels from certain interesting angle or levels of approximations but is also inevitably subject to its limitations and applicability. On the other hand, at some deeper levels, they share common grounds and roots, from which more efficient hybrid tools or methods can be developed. This highlights the necessity of integrating this diversity of approaches.

4.1 Noise

Because of random processes linked to image formation, acquisition, recording and transmission, images become *always* corrupted with noise, which manifests itself as stochastic variations of image intensities. This random nature differentiates it from deterministic interferences, shading, lack of focus, and many other distortions. In some cases, small amounts of high-frequency noise add to the "naturalness" of textures (in contrast with a plastic or synthetic appearance) and have been found to increase perceived quality [75]. Most often, however, the presence of noise in an image degrades both, the perceived quality (giving the subjective impression that the image is not clean, or that something unintended is superimposed on the image), as well as the performance of the task for which the image is intended (e.g., medical diagnosis), if not defeat it altogether, since it limits the amount of the visual information.

The presence of noise in images is unavoidable due to its inherent nature and its statistical, random characteristic. Even a *professional* photo is bound to have some noise in it. Therefore, effective ***denoising*** (i.e., the process of estimating the original image information from the observed noisy data) is an essential part of many image processing systems as a pre-processing for other image tasks, e.g. compression, segmentation and recognition, in order to produce reliable results in various image-related applications, such as aerospace, medical image analysis or object detection[66].

As such, image denoising is perhaps the "simplest", and at the same time one of the most important of image processing problems that image processing and computer vision researchers have studied over a long time. Nevertheless, it remains a wide-open field of research as we see progress in image capturing sensor technology, since the increase in the number of pixels in sensors typically translates to smaller pixels which consequently gives rise to an increase in the perceived noise in the captured image due to the availability of fewer photons. This makes image denoising still an even more relevant problem necessitating continuing research. In the recent years various new techniques have been proposed which perform very good denoising even in the presence of large amounts of noise. Among the most appealing aspects of this field are the ability to refer it to a well-established theory, and the fact that the proposed algorithms in this field are efficient and practical [75].

While this chapter focuses on smoothing for the specific application of image denoising, the tools developed and the results obtained can easily be extended to other application areas where it is required to split an image into its *intrinsic components*, such separation of reflectance and illuminance in tone management, as shown in next chapter.

---

[66] Image *deblurring* and *denoising*, which respectively deal with light measure uncertainty in the spatial and range domains, are among the most fundamental problems in image processing.

4.1.1 Image Quality and Noise

Although noise may also affect other attributes such as sharpness, the perceptual attribute of image quality that is most strongly influenced by noise in an image is the *perceived noise* or *noisiness* (also referred to as *visibility* or *annoyance* of noise), which has itself been identified to be one of the most important dimensions of image quality. Until now the most current method to quantify noise has been the signal to noise ratio (SNR) measured with respect to a true image (often a uniform patch, sometimes an edge or an oscillation[67]). It describes the behavior of noise in digital image capture devices, but this does not always match the perception of noise, *noisiness*, as it is widely acknowledged. Being able to evaluate the noise using a quantization based on the perception has become necessary, not only for IQ assessment, but also for evaluating the performance of denoising algorithms.

Many researchers have studied the effect of noise on image quality in psychophysical experiments, e.g., in a seminal paper Dooley and Shaw [85] proposed a metric that integrated the Weiner noise power spectrum with properties of the human visual system, specifically the contrast sensitivity function. This and many similar approaches have shown a degree of success in regard to predicting the noisiness for uniform patches: specifically, their studies indicate that, while the noisiness of an image may depend on many parameters of both noise and image, the most influential ones are the noise ***standard deviation*** (SD) and ***correlation length*** (CL), in regard to respectively the *strength* and *size*, *spread* or *bandwidth* of non-white noise. In uniform regions of the image, where it is most noticeable, the mean local luminance may also affect the noisiness. However, it is found to be roughly independent of the probability density function (PDF) of the noise. Besides, for colour images overall the luminance noise was most perceptible, resulting in the biggest decrease in quality, while the high-frequency chromatic noise was judged to have the small effect on quality, suggesting it was almost imperceptible. Based on these findings an objective measure of the noisiness of an image, i.e., *noise index*, can be computed [95]. However, such systematic approaches cannot address the effect of noise perception in existing complex images, as they all require the measurement of patch data. In order to understand the relationship these systematic metrics have with the perception of overall image quality, a better understanding of noise in complex images is required.

This thesis takes in turn an innovative approach, based on a pure functional objective relation between noise and IQ. According to the information-based approach to IQ presented in Chapter 3, noise reduces image quality because, due to its inherent random nature, introduces uncertainty in the range domain, which limits the amount of information that can be extracted from the image[68].

---

[67] Standard practice is described in ISO15739:2003 [123].

[68] Recall that the quality of something was defined somewhere as how well it serves the purpose for which is intended. Images are, above all, carriers of visual information.

Accordingly, an appropriate way of characterizing the effect of noise is by means of its *entropy*. For a white Gaussian noise (which, for a given variance, is the distribution with highest entropy), the (differential) entropy is expressed as

$$H(dB) = 1/2 \log\left\{(2\pi)^N |\Sigma|\right\} \quad IQ \propto SNR(dB) = 1 - 2H$$

where $|\sum|$ is the determinant of the covariance matrix (i.e. the noise power). Then, the quality of a noisy image is evaluated by means of its SNR(db), which correlates with perceived degradation much better than the MSE (see Chapter 3), as given by eq. 2.[69]

In order to *i)* better describe the perception of the noise; *ii)* automate subsequent IP steps; and *iii)* be able to predict noise after various treatments occurring in imaging chains before the actual printing or viewing of the image, we propose to extend the standard above SNR-based framework with the correlation function of the noise, $\boldsymbol{R_n(x)}$, which gives rise to interesting figures of merit such as the quadratic size/spread, $\boldsymbol{\rho^2}$, and the invariant noise level, $\boldsymbol{\lambda}$

$$\rho^2 = \int x^2 R_n(x)dx \Big/ \int R_n(x)dx \qquad \lambda = \int R_n(x)dx$$

In section 4.5 the noise estimators for white and non-white noise are derived and their performance discussed.

See *Sources of noise* (Appendix)

4.1.2 Generalized Signal-Dependent additive noise model

While most of algorithms found in the literature assumed without justification a signal-independent additive white Gaussian noise (a.k.a AWGN), noise in real-world digital images exhibits significant dependency on local signal (see Appendix). It may also happen that the noise is independent from the signal, but not additive (e.g. multiplicative). In all these cases, the noise may be interpreted as signal-dependent, under an additive-noise (Gaussian) model [83]. Another possibility is that noise statistics depend on the spatial location [110]. For all these situations, it is convenient to refine the additive independent global Gaussian noise model, to allow for spatially varying noise statistics. Besides signal dependency, spatial and cross-channel correlation are also important features of the noise arising in many capture devices. Therefore, we can only expect high-performance noise estimation and denoising with real images after having accounted for all these features. In what follows a more realistic CCD noise model than the classical one is presented, leading to the basic hypothesis justifying the filter election.

[69] Following this approach, blur in turn introduces uncertainty in the spatial domain and hence is modeled in the information-based approach as a signal attenuation (think of it in Fourier domain).

4.1.2.1 Signal Dependency

Let ***u(i) [0,1]***, the normalized scene irradiance at pixel ***i***, represent the ***"true"*** light intensity average power sent by the scene to pixel ***i***, normalized by the maximum measurable value [70]

$$u(i) = \frac{E\left[N_I(i)\right]}{C} = \frac{E\left\lfloor N_g(i)\right\rfloor}{L} = p$$

Let ***f*** denote the camera response function (CRF)[71], which translates the ***measured*** scene irradiance at pixel ***i*** to image brightness ***v(i) [0,1]***

$$v(i) = f\left(u(i) + n_s(i) + n_c(i)\right) + n_q(i)$$

where $n_s(u(i))$, the signal-dependent noise term, accounts for *photographic* and *photon* noises; n_c, the signal-independent noise term, accounts for *thermal* and *read-out* noises; and n_q is the *quantization* noise. By now, we will ignore n_q and expand v(i) as follows, what yields an additive noise model

$$\begin{aligned} v(i) &= f\left(u(i)\right) + f'\left(u(i)\right)\cdot\left(n_s(i) + n_c(i)\right) \\ &:= f\left(u(i)\right) + n_0(i) \end{aligned}$$

There are two different regimes in which CCD imaging is used. In low light dominate the thermal and read-out noises, while in high light levels (which is the regime in which consumer imaging devices are normally used) ***u(i)*** dominates the noise terms, among which the most important is the shot noise (see [75]. Ch 4.5). For large values of N_I, the Poisson distribution is well-modelled as Gaussian and the overall noise may be interpreted as signal-dependent *additive white Gaussian noise* (AWGN)

$$n_0(i) \approx N(0, {\sigma_0}^2(i)), \text{ where}$$

$${\sigma_0}^2(i) = \begin{cases} f'\left(u(i)\right)^2 \cdot (\sigma_{th}^2 + \sigma_{ro}^2) & , \ u(i) \ \ low \\ f'\left(u(i)\right)^2 \cdot \left[\left(\underbrace{u(i)\left(1-u(i)\right)/L}_{photographic\ noise}\right) + \underbrace{u(i)/C}_{shot\ noise}\right] & , \ u(i) \ \ high \end{cases}$$

where both, binomial and Poisson distributions (corresponding to photographic and shot noises, respectively) have been approximated by Gaussian distributions.

Observe that the noise model at each pixel ***i*** is white (i.e., not correlated) and only depends on the original pixel value ***u(i)*** and is additive. Thus, the overall noise looks Gaussian, but the signal to noise ratio is higher in bright regions than

---

[70] Normalization allows us to forget units, unifying notation for different capturing systems (e.g. CCD vs. film), independency of bit-depth and concentrating on noise. It is surprising that many authors still give absolute instead of relative noise amounts figures.

[71] In practice, ***f*** can be approximated by the gamma correction function, i.e., $f(x) \approx x^{1/\gamma}$, $0<1/\gamma<1$.

in dark regions. For a 2nd order characterization of the noise, mean and variance are enough. If noise is Gaussian, then its pdf is completely characterized.

In addition, one may substitute the assumed Gaussian distribution by a generalized Gaussian, which has the advantage of being able to fit a large variety of (symmetric) noises by appropriate choice of the three parameters μ, σ, and α [75]

$$p(x) = Ae^{-\beta|x-\mu|^{\alpha}}, \text{ where}$$

$$\beta = \frac{1}{\sigma}\left(\frac{\Gamma(3/\alpha)}{\Gamma(1+\alpha)}\right)^{0.5};$$

$$A = \frac{\alpha\beta}{2\Gamma(1/\alpha)}$$

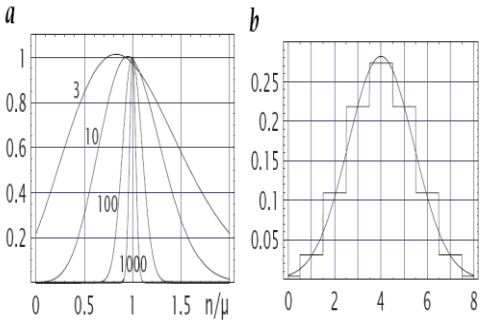


Figure 4.1: Approximation of binomial and Poisson by Gaussian when large number (a); and SNR improvement as relative uncertainty (measured here by the normalized std.) decrease with N of a Poisson process (b).

4.1.2.2 Intra- and Cross-channel Correlation

In the previous model the noise is signal dependent but independent at different pixels, yielding a flat power spectrum. This is however a fairly unrealistic assumption in most practical cases, where noise often presents a low-pass (i.e. blurred-like) behaviour. In the case of non-white (i.e., correlated) noise, variance must be replaced by a variance-covariance matrix. We will assume Gaussian noise and independency between spatial and range domains, so that the variance-covariance matrix can be factorized into two matrixes which respectively deal with intra- and inter- frame noise correlations. We will also assume that the auto-covariance function of the non-white noise is Gaussian, and hence characterized by two parameters, the noise standard deviation and the noise correlation length. [95] have develop an algorithm for simultaneously estimating both parameters by analysing the image at two scales.

Besides, in colour imaging we may also find noise with cross-channel correlations introduced by transformation between colour spaces. To account for these spatial and cross-channel correlations, we define

$$\mathbf{n} = \mathbf{n_1}\mathbf{R_{CTM}} \qquad\qquad n_1 = n_0 * \bar{k}$$

where $\bar{k} = k/\|k\|_{L_2}$, a normalized convolution kernel, accounts for intra-channel spatial correlations; and $\boldsymbol{R}_{CTM}$, a BxB correlation matrix, accounts for the colour or inter-channel correlations resulting from applying a colour transformation matrix. Note that we have now used a vectorial notation to remark that at each pixel we have not a single value, but a vector (with as many components as bands or channels the image has).

4.2 Noise Reduction: Theoretical Framework

The image denoising problem is important, not only because of the evident applications has it served as an important part to digital image acquiring systems to enhance image quality. Being the simplest possible inverse problem, it provides a convenient platform to examine natural image models and signal separation algorithms over which image processing ideas and techniques can be assessed.

For image denoising we assume that the original image is degraded only by the presence of noise. The introduction of noise into a signal is often modelled as an additive process. Then, the observation model for each pixel i in the image can be mathematically written as $v_i = u(x_i) + n_i$, where $u(\cdot)$ denotes a real function describing the "true" (unknown) pixel value which is corrupted by additive (unknown) noise n_i resulting in the observed value v_i at location $x_i = [x_1\ x_2]_i^T$. The goal of denoising is then to remove (or at least minimize) the noise effectively while at the same time preserving important features, such as edges and fine details, as much as possible. In other words, we desire to design an algorithm that can remove the noise from v, getting as close as possible to the original image, u. The most common performance criterion is the mean squared error (MSE), $MSE(u, \hat{u}) = E(u - \hat{u})^2 = ||u-\hat{u}||^2$. Formally, given a noisy image v, we wish to compute an estimate of the original (clean) image $\hat{u} = f(v)$, where the estimator f is selected from a family F to minimize the MSE

$$f_{opt} = f_{MMSE} = \arg\min_{f \in F} E\{|u - f(v)|^2\}$$

and $E\{\cdot\}$ indicates the expected value. Some authors refer to this as *filtering* problem, while others say it is an *estimation problem*. In either case, u is respectively regarded to be fixed but unknown (the so called *"frequentist"* perspective), or a sample drawn from some prior probability distribution $p_u(U)$ (the *Bayesian* perspective).

The quality of a corrupted image is evaluated by using the *peak signal to noise ratio* (PSNR) measured in dB: $PSNR_{dB} = 10\ log_{10}\ (1/MSE)$. A *filtering gain* or *improvement signal-to-noise ratio* (ISNR) is then defined as a metric to quantify performances: $ISNR = PSNR_{out} - PSNR_{in} = 10\ log_{10}\ (||u-v||^2/||u-\hat{u}||^2)$. Observe that, while the MSE is not computable in a real problem and its results are not well correlated with the HVS, it has the advantages of easy tractability and intuitive appeal since MSE can be interpreted as '*noise power*'.

Most existing denoising methods can be basically divided into two categories: *image or spatial domain* filtering methods and *transform domain* filtering methods. The former operates in a single resolution and performs averaging of neighbouring pixels to achieve noise smoothing, while the latter performs decomposition of a signal into sub-bands in order to apply some kind of coefficient shrinkage and then inverts the transform. Such methods try to employ the energy compaction property of different transforms (e.g. discrete wavelet transform -DWT- or Karhunen-Loeve transform) to better separate image data

from noise data. The most widely used family of methods of this type is *wavelet thresholding*. Wavelets have a property of shape invariance of basis functions which allows one to control the Gibbs phenomenon via careful selection of a wavelet basis. Algorithms in spatial domain filtering can be further categorized according to *linear* vs *non-linear*, *local* vs *non-local*, *variational* vs. *statistical*, etc. In this chapter, we have no intention to perform a deep survey of this vast activity, but just to provide a proper framework.

Weighted Least Squares (WLS): denoising by averaging

The problem of recovering u from v is an ill-conditioned inverse problem in that knowledge of the direct problem (i.e., information provided by v and the observation model) is not sufficient to ensure the existence, uniqueness, and stability of a solution $\hat{u}$. Image denoising methods use regularization techniques based on a priori knowledge of the image properties to approximate u, such as the degree of smoothness, total variation, decay as well as sparsity of the transform domain coefficients, etc., which have been exploited in various image processing tasks, including compression [75].[72] Many of these regularity properties are *local*, in the sense that the greyscale value at a pixel is correlated with values in its neighbourhood. Self-similarity is an example of a *non-local* regularity property, in the sense that local neighbourhoods of an image can be highly correlated (i.e., affinely similar) to other neighbourhoods in the image.

Among the numerous methods to suppress noise, we are focusing on the family of methods based on Maximum A-posteriori Probability (Bayesian) Estimation, which allows consistently to combine empirical data and prior knowledge about the image, such as piece-wise smoothness assumption

$$\hat{u} = \arg\min_{u} \{ \underbrace{(v-u)^2}_{data\_fidelity\_term} + \underbrace{\lambda R(u)}_{prior\_smoothness\_term} \}$$

State-of-the-art methods from this family are: *Anisotropic diffusion*, *Weighted Least Squares*, *Robust Estimators*, and the *Bilateral filter*. Even though they may be very different in tools it must be emphasized that a wide class, regardless of implementation, share the same basic idea: denoising is achieved by averaging (either pixel intensities in the spatial domain or coefficient values in a transform domain). The variance law in probability theory ensures that if nine pixels with the same colour plus some decorrelated noise are averaged, then the noise in the average is divided by three.[73]

[72] A simple and well-known regularization supposes that images are globally smooth, enforcing a roughness penalty on the solution. This a priori constraint is very important for separating noise from a clean image, or reflectance from illuminance (i.e., for those applications requiring splitting the input image into two layers: a large-scale component, which is a smoothed version of the input, and a small-scale component, which is the residual of the filter).

[73] Denoising by linear averaging is grounded on a second order statistical characterization of image and noise. If these differ, then they can be separated by a decorrelation transform (Fourier, PCA, etc.)

$$\hat{u}(i) = \frac{1}{C(i)} \sum_{j \in I} w(i,j) v(j) \;\Leftrightarrow \qquad \hat{u}_i = \frac{1}{C_i} \sum_{j \in I} w_{i,j} v_j \text{ , where } C_i = \sum_{j \in I} w_{i,j}$$

where C_i acts as a normalizing constant to ensure that a constant function $\underline{v_j}$ is mapped to itself. This can be recognized as a *weighted average* or *weighted least squares* (WLS) (method).

The performance of this approach depends on the filter parameters (weights) $w_{i,j}$, and several ways to compute them have been suggested in the literature [75],[101],[115],[116],[118]. While in linear filtering the weights are fixed and do not depend on local context, in the more general framework of adaptive filtering the weight function is defined on the basis of the local context.

Window-based Filtering

Filtering is perhaps the most fundamental operation of image processing and computer vision. In the broadest sense of the term *"filtering"*, the value of the filtered image at a given location is a function of the values of the input image in a small neighbourhood of the same location (namely, a sliding window will pass over the image to capture information in a localized area that will be used to determine the output value of the pixel at the centre of the window). This approach is chosen due to the localized nature of image features and the point-wise model of noise formation, and is common to all the applications described here.[74] The following will discuss the specific filtering frameworks.

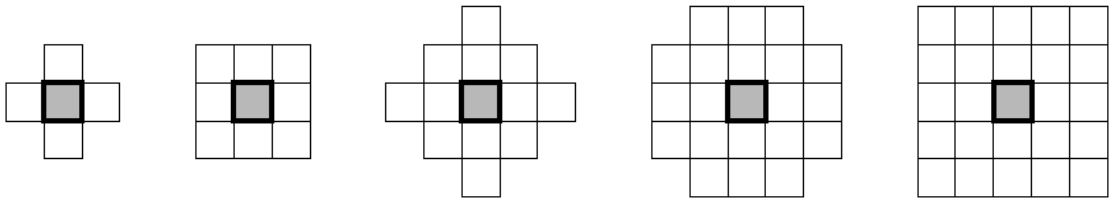


Figure 4.2: Spatial windows, a.k.a. *neighbourhoods*. The union of the pixel being processed and its neighbouring pixels is commonly referred to as *window*, a *mask*, or *neighbourhood*. Local windows typically involve fewer than 7x7 pixels, on images with up to 10^7 pixels. The unbiased window configuration shown in the figure is known as *isotropic*.

Assuming that the statistics are spatial-invariant (Markovian assumption) this leads to spatially invariant smoothing, resulting in constant (and thus non-adaptive) averaging window operators (e.g., low-pass filtering using Gaussian kernels). Indeed, it is a common practice in computer vision and image processing to convolve rectangular fixed windows with digital images to perform local smoothing. If all the pixels in the window come from the same population as the central pixel, this practice is reasonable and fast. But, as is well

[74] Although formal and quantitative explanations of this weight fall-off can be given, the intuition is that images typically vary slowly over space, so near pixels are likely to have similar values, and it is therefore appropriate to average them together. The noise values that corrupt these nearby pixels are mutually less correlated than the signal values, so noise is averaged away while signal is preserved.

known, constant coefficient window operators produce incorrect results if more than one statistical population is present within a window, e.g., when it overlaps a discontinuity, since it results in averaging information from different regions near edges, which are consequently blurred by low-pass filtering.

A more realistic image model assumes that images are made of smooth regions, separated by sharp edges or boundaries (known as *piece-wise* smoothness assumption). This is called *edge-preserving* regularization or smoothing and requires higher order characterization, more localized transforms (wavelets) and/or varying weights depending upon local image structure, leading to nonlinear algorithms, more effective than linear ones, but also more difficult to analyse. In order to avoid removing important image features, pixels may be averaged using space dependent kernels, having their size, shape and weight coefficients adapted to the local image structure.

Example-based techniques. Statistical neighbourhood approaches

In addressing the general inverse problems in image processing using the (Bayesian) classical approach, an image prior is necessary. Traditionally, this has been handled by choosing a prior based on some simplifying assumptions, such as spatial smoothness, low/max-entropy, or sparsity in some transform domain. While these common approaches lean on a guess of a mathematical expression for the image prior, the methods here presented suggest that it is possible to take advantage of an image model learned from the observed image itself. More specifically, these denoising methods attempt to learn the statistical relationship between the image values in a window around a pixel and the pixel value at the window center:

u_i as a function of its position $\boldsymbol{x}_i$: $u_i = u(\boldsymbol{x}_i)$, where $\boldsymbol{x}_i = [x_1,x_2]_i^T$ -> $p(u_i \mid x_i)$

u_i as a function of its neighbors $u_{\beta(i)}$ distribution: $u_i = u(u_{\beta(i)})$ -> $p(u_i \mid p(u_{\beta(i)}))$

u_i as a function of its neighborhood: $u_i = u(\mathcal{N}_i)$ -> $p(u_i \mid \boldsymbol{N}_i)$

where the only assumption is simply that the function describing such statistical relationship (namely the *regression function*) is smooth as a function on respectively the image space, the space of intensities and the space of patches.

This provides a common framework for conventional low-pass filters as well as bilateral filters, since both can be described as special cases of the proposed method.

In the last decade, several concepts related to the general theory we promote here have been rediscovered in different guises, and presented under different names such as *normalized convolution* [99], *bilateral filter* [116], *edge-directed interpolation*, and *moving least squares*. However, it is still not easy to see the advantages of the various approaches, and the relations between different methods are only partly understood. This rest of this chapter is intended as a

contribution in this direction: by studying several methods and their relations, we end up with a better understanding of each of them.

4.2.1 Smoothness-based methods

The methods presented from the diffusion framework consider images in a continuum and regard noise reduction as an evolution process characterized of local pixel interactions described by partial differential equations [75], [108].

Regularizing an image **u** may be seen as the minimization of a functional $E_s(u)$ measuring a global spatial (image) variation, since this will enforce smoothness, removing the noise gradually. To that end, the simplest choice as difference operator containing only first-order derivatives, is the gradient ∇ operator [75]

$$E_S(u) = \int_{\Omega} |\nabla u|^2 \, dx \qquad (4.1)$$

When gradient descent methods are applied to variational problems like this one, they frequently give rise to PDE-based smoothing methods, among which the simplest and best investigated ones are those based on a *diffusion process*, "*a physical process that equilibrates concentration differences* (expressed by Fick's law*) without creating or destroying 'mass'* (expressed by the continuity equation)"

$$j = -\mathbf{D}\nabla u \quad (4.2) \qquad \frac{\partial u}{\partial t} = -div(j) \quad (4.3) \qquad \frac{\partial u}{\partial t} = div(\mathbf{D}\nabla u) = \mathbf{D}\Delta u \quad (4.4)$$

where u is the concentration, j is a flux which aims to compensate for the concentration gradient ∇u, and the diffusion tensor **D** is a positive definite symmetric matrix.

Equation (4.4) appears in many physical transport processes. In the context of heat transfer it is called *heat flow equation*. In image processing, we may interpret the concentration u as a gray value at a point (pixel), resulting in three relevant cases:

a) linear isotropic diffusion filters: $\mathbf{D}=d\mathbf{I}$ (where d is the *diffusivity*)

b) nonlinear isotropic diffusion filters: $\mathbf{D}(x)=d(x)\mathbf{I}$ *(d(x)*: diffusivity depends on spatial position/location, i.e., it is not spatially invariant)

c) nonlinear anisotropic diffusion filters ($\mathbf{D}(x) \neq$ *diagonal*) with diffusion tensor depending on the local image structure [108].[76]

Linear isotropic diffusion is equivalent to a Gaussian convolution, a low-pass filter. In anisotropic diffusion, the conductance depends on the image, and both the image and the conductance evolve over time in more interesting ways. We present here the linear case and treat the nonlinear ones in section 4.3.1.

---

[75] Observe that, since the intensity function of a digital image is only known at discrete points, derivates of this function cannot be defined unless we assume that there is an underlying continuous intensity function which has been sampled at the image points.

[76] This is done via the so-called ***structure tensor*** (second moment matrix), a well-established tool for texture analysis.

Minimization of (4.1) subject to noise constraints (4.5) and (4.6) defines the constrained optimization problem in (4.7) [77]

$$E[n] = 0 \Rightarrow \int_\Omega u dx = \int_\Omega v dx \qquad (4.5) \qquad E[n^2] = \sigma_n^2 \Rightarrow \int_\Omega (u-v)^2 dx = \sigma_n^2 \qquad (4.6)$$

$$\mathrm{E}(u) = \mathrm{E_D}(u) + \lambda \mathrm{E_S}(u) = \int_\Omega \left(|u - v|^2 + \lambda_S |\nabla u|^2 \right) dx \qquad (4.7)$$

Note that every minimizer of eq. (4.7) has to necessarily satisfy the Euler-Lagrange equation (4.8). Its solution can be regarded as the steady state of the *diffusion–reaction*[78] process in (4.9) [75], and yields the filtered image at infinite time[79]. Rewriting (4.8) as in (4.10) it becomes evident that this process can be regarded as an implicit time discretization of the diffusion process as described in (4.4), with single time step of size λ, and $u_i^0 = v_i$ in order to recover the similarity constraint. It gives the desired result at finite diffusion time $t=\lambda$.

$$0 = \Delta u - \frac{u-v}{\lambda} \qquad (4.8) \qquad \frac{\partial u}{\partial t} = \Delta u - \frac{u-v}{\lambda} \qquad (4.9)$$

$$\frac{u-v}{\lambda} = \Delta u \qquad (4.10) \qquad \boxed{u_i^{k+1} = u_i^k + \frac{\lambda}{4} \sum_{j \in \beta_4(i)} (u_j^k - u_i^k)} \qquad (4.11)$$

Given a discrete sampled image u, equation (4.10) can be spatially discretized on a square lattice, with brightness values associated to the vertices and conduction coefficients to the arcs. A 0th order approximation of the Laplacian yields to the iterative formulation in (4.11), where k denotes discrete time steps, $\lambda \in \Re^+$ determines the rate of diffusion, $\beta_4(i)$ denotes the spatial neighborhood of pixel i formed by its four nearest neighbor.

Observe that *(i)* the constraint given by (4.6) (i.e., the data term) is implicit in the initial condition $u_i^0 = v_i$, and *(ii)* diffusion needs to be stopped in order not to get a completely flat image. This connects to the idea of images as manifolds evolving to minimal surfaces.

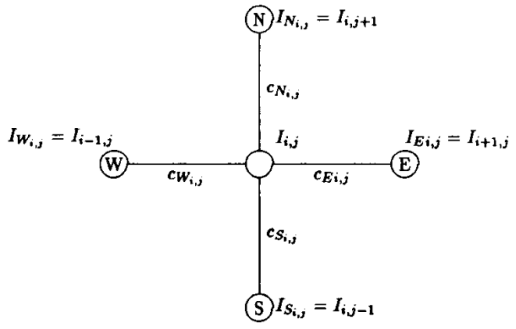


Figure 4.3: The structure of the discrete computational scheme for simulating the diffusion equation. The brightness values Ii,j are associated with the nodes of a lattice, the conduction coefficients c to the arcs [108].

[77] Because of the translation invariance of $E_s(u)$ (i.e., $E_s(u) = E_s(u+c)$, for any constant c), the constraint (4.5) is in fact already encoded.

[78] Observe that the second term, without which u would slowly smooth out until becoming flat, tends to pull u back toward the observed image v.

[79] Here the "time" t is a purely numerical parameter.

4.2.2 Data similarity-based methods

In contrast to smoothness-based variational methods, which lead to algorithms based on PDE's, data-term minimization subject to some smoothness constraint leads to constant coefficient window operators, for which local surface fitting based on nonparametric regression provides the theoretical basis.

The kernel regression framework defines the observation model as

$$\mathbf{v}_i = u(\mathbf{x}_i) + \mathbf{n}_i \qquad (4.12) \qquad\qquad u : \Omega \to \Re \qquad (4.13)$$

where v_i is the observed noisy sample at x_i, $u(x)$ is the (unspecified) ***regression function*** to be estimated and n_i is an i.i.d. zero mean noise.[80]

Although the specific form of $u(x)$ may remain unknown, assuming that the underlying image data is locally smooth to some order M, we can rely on a Taylor[81] expansion about x_i of the form

$$\begin{aligned} u(\mathbf{x}_j) &= u(\mathbf{x}_i) + \{\nabla u(\mathbf{x}_i)\}^T (\mathbf{x}_j - \mathbf{x}_i) + \frac{1}{2!}(\mathbf{x}_j - \mathbf{x}_i)^T \{\mathbf{H} u(\mathbf{x}_i)\}(\mathbf{x}_j - \mathbf{x}_i) + \ldots \\ &\approx \sum_{k=0}^{M} \beta_{\{k\}_i} (\mathbf{x}_j - \mathbf{x}_i)^k \end{aligned} \qquad (4.14)$$

where H denotes the Hessian operator.

The optimization problem in eq. 4.7 may now be written as a data-term minimization, subject to the implicit smoothness constraint derived from considering just the first M+1 terms in the previous expansion of u(x)

$$\min_{\mathbf{u}} \mathrm{E}(\mathbf{u}) \Leftrightarrow \min_{\mathbf{u}} \mathrm{E}_{\mathrm{D}}(\mathbf{u})\Big|_{u(\mathbf{x}_j)=\sum_{k=0}^{M} \beta_{\{k\}_i} (\mathbf{x}_j - \mathbf{x}_i)^k} \qquad (4.15)$$

Here each element β_i is to be estimated from the observed data, typically using a (locally) weighted least squares approach minimizing the sum of the squared residual errors $e_i = v_i - u(x_i)$

$$\begin{aligned} \hat{\beta}_i &= \arg\min_{\beta_i} \sum_{j \in \beta(i)} | \mathbf{v}_j - \sum_{k=0}^{M} \beta_{\{k\}_i} (\mathbf{x}_j - \mathbf{x}_i)^k |^2 \\ &= \arg\min_{\beta_i} \sum_{j=1}^{N} | \mathbf{v}_j - \sum_{k=0}^{M} \beta_{\{k\}_i} (\mathbf{x}_j - \mathbf{x}_i)^k |^2 \; K_{\mathbf{H}}(x_j - x_i) \end{aligned} \qquad (4.16)$$

Observe that the estimator of u(xi) is $\hat{\beta}_{\{0\}i}$, and $K_H(\cdot)$ is a so-called ***kernel function*** that penalizes distance away from the local position x_i where the approximation is centered, and H is a $2x2$ ***smoothing matrix***.

---

[80] While there are several other effective nonparametric regression methods such as spline interpolation, orthogonal series or local polynomial, the kernel regression framework provides a rich mechanism for computing point-wise estimates of the regression function with minimal assumptions about global signal or noise models [115].

[81] Other localized representations are also possible and may be advantageous.

$$K_{\mathbf{H}}(\mathbf{t}) = \frac{1}{\det(\mathbf{H})} K(\mathbf{H}^{-1}\mathbf{t}) \quad (4.17) \qquad |\mathbf{t}|^2_{\mathbf{H}} = \mathbf{t}^T(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{t} \quad (4.18)$$

While higher order approximations are also possible, locally constant, linear and quadratic (corresponding to *M = 0, 1, 2* respectively) have been considered most widely in the regression literature. We will concentrate on the simplest case (*M=0*), assuming that *u(x)* is a locally constant function (i.e. $u_i = u_j$ if $j \in \beta_r(i) \Leftrightarrow \| \mathbf{x}_i - \mathbf{x}_j \|^2 < r$), obtaining

$$\hat{u}_j = \frac{\sum_{i \in L} K_{\mathbf{H}}(x_j - x_i) v_i}{\sum_{i \in L} K_{\mathbf{H}}(x_j - x_i)} \quad (4.19)$$

which is the well-known ***Nadaraya-Watson Estimator*** (NWE), introduced in the statistical literature more than 40 years ago.[82] This is an estimator of the conditional expectation $\hat{u}(x) = E[\phi(V) \mid X=x]$. In our case, $\phi(V)=V$.

4.2.2.1 Convolution Kernel election

We are interested mostly in a special class of radially symmetric, also known as *isotropic*, kernels satisfying

$$K_H(v - v_i) = k(\| v - v_i \|^2_H)$$

where the weighted norm $\| v - v_i \|^2_H = (v - v_i)^T H^{-1}(v - v_i)$ is the ***Mahalanobis distance*** from v to v_i. Often the profile $k_N(x) = \exp(-1/2 \cdot x)$ ($x \geq 0$) is chosen, resulting in the multivariate kernel $K_N(x) = (2\pi)^{-d/2} e^{-1/2\|x\|^2}$, which is symmetrically truncated to have a kernel with finite support.

4.2.2.2 Equivalence of Gaussian convolution and linear diffusion

Remark that the optical blur is equivalent to one step of the heat equation. In fact, it can be shown that for any bounded $v \in C(\Re^2)$, the linear diffusion process possesses the unique solution[83]

$$\frac{\partial \mathbf{u}}{\partial t} = div(\nabla \mathbf{u}), \ \mathbf{u}(x,0) = \mathbf{v}(x) \quad (4.20) \qquad u(x,t) = \begin{cases} v(x), t = 0 \\ \left(K_{\sqrt{2t}} * v\right)(x), t > 0 \end{cases} \quad (4.21)$$

Hence, smoothing structures of order σ requires to stop the diffusion process at time $T = 1/2\sigma^2$. Alternatively, it is easy to prove that $K_\sigma = K_{\sigma/2} * K_{\sigma/2}$, viewing each step of linear diffusion as convolving with K_σ ($\sigma \to 0$ when $\lambda \to 0$). Since the optical blur is equivalent to one step of the heat equation, we can, to some extent, deblur an image by reversing the time.

---

[82] This is often described as a kernel operation. The use of weighting kernels to average together pixels in a neighborhood is a convolution operation.

[83] This solution is unique, provided we restrict ourselves to functions satisfying $|u(x,t)| \leq M \cdot \exp(a \mid x \mid^2)$, $(M, a > 0)$.

4.3 Edge-preserving smoothing

While leading to satisfactory solutions of some of early vision problems, standard regularization makes strong geometric assumptions to impose regularity on the original image, smoothing out details and fine structures because they behave in all functional aspects as noise [84]. ***Edge-preserving filters*** smooth images without disturbing the sharpness and position of edges, which may carry a considerable amount of information in an image, playing a critical role in our perception as well as in the analysis of images, a fact that becomes particularly evident in line drawings or hand-writing. To avoid diffusion (averaging) across edges while at the same time keeping many averaged values, smoothness- and data-based algorithms must be done adaptive to local image structure, resulting in nonlinear filters. How to control the balance between *variance* and *bias* according to image's local features is a key problem in adaptive signal processing. [85]

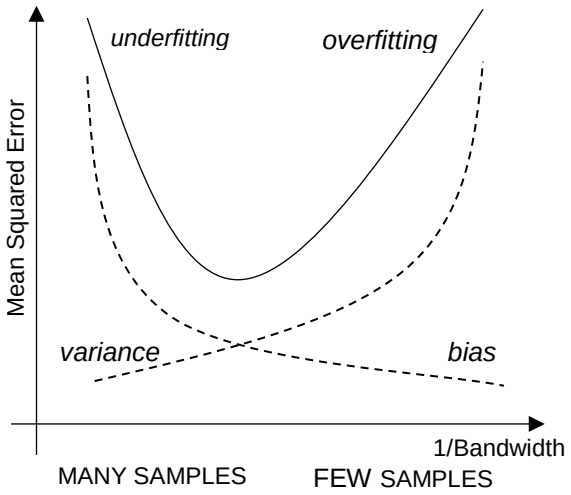


Figure 4.4: Bias-variance tradeoff as a function of the number of averaged samples (*kernel's bandwidth*).

Edge-preserving smoothing is closely related to boundary detection, which is an own discipline in imaging. Instead of complex modeling of the prior information to explicitly incorporate structures like edges, lines, corners, or even texture about the image to be recovered, in this section, we study three different frameworks (*heuristic* nonlinear improvements, *stochastic* regularization and *robust* regularization) where discontinuities are addressed implicitly rather than explicitly, regarding them as outliers of the local model. This is achieved either ***i)*** reducing filtering effect (decrease bandwidth) where an edge is present (*isotropic* kernel), ***ii)*** changing shape but not size (*anisotropic* kernel), and ***iii)*** by weighted average according to likelihood of belonging to the same original class as the central pixel. While many filters have been proposed, not yet a unifying framework has been provided.

---

[84] Notice that standard regularization theory with linear A and P is equivalent to restricting the space of the solution to generalized splines, whose order depends on the stabilizer P.

[85] Much insight into the performance of nonparametric regression estimators $\hat{m}$ has come from studying the mean squared error (*MSE*), which may be separated into a variance term and a bias term: $MSE(x) = E(\hat{m}(x) - m(x))^2 = \text{var}(\hat{m}(x)) + (E(\hat{m}(x)) - m(x))^2$ .The first term is the variance of the estimator, $\hat{m}$ and will decrease as the bandwidth is increased, since more data will be averaged, giving a more reliable estimate. The second term is the squared bias, which will conversely increase as the bandwidth is increased, since the fit will in general become worse, which means that the estimate $\hat{m}$ will not give a good estimate of the true value of m . Thus, we have a bias-variance tradeoff.

4.3.1 Heuristic nonlinear improvements to standard regularization

As classical regularization, these methods were proposed within the framework of geometrical modeling of images.

4.3.1.1 Nonlinear anisotropic diffusion

In their seminal paper, Perona and Malik noted in [108] that conductance controls the rate of local image smoothing and proposed an *edge-stopping* function *g* that varies the conductance inversely with a local "edginess" estimate, in order to find, preserve and sharpen image edges. They used gradient magnitude scalar $|\nabla \mathbf{u}|$ for the edginess estimate

$$\frac{\partial \mathbf{u}}{\partial t} = div(\mathbf{D}\nabla \mathbf{u}), \text{ where } \mathbf{D} = g\left(|\nabla \mathbf{u}|^2\right) \qquad (4.22)$$

and proposed two expressions for the edge-stopping function g(s²)

$$g_1(s^2) = \frac{1}{1+s^2/\sigma^2} \qquad (4.23) \qquad g_2(s^2) = e^{-(s^2/\sigma^2)} \qquad (4.24)$$

where σ is a scale parameter in the intensity domain that specifies what gradient intensity should stop diffusion. Equation (4.11) now results in the nonlinear filter

$$u_i^{k+1} = u_i^k + \frac{\lambda}{4}\sum_{j\in\beta_4(i)} g_S(|u_j^k - u_i^k|^2)(u_j^k - u_i^k) \qquad (4.25)$$

4.3.1.2 Extended support diffusion: linear Gaussian smoothing

The number of iterations required can be reduced by simply extending the 0th order diffusion of eq. to a larger spatial support [79]

$$\begin{aligned} u_i^{k+1} &= u_i^k + \frac{\lambda}{4}\sum_{j\in\beta_4(i)} g_S(|u_j^k - u_i^k|^2)(u_j^k - u_i^k) \\ &= u_i^k + \lambda\sum_{j\in\beta(i)} g_S(|u_j^k - u_i^k|^2)w(|x_j - x_i|^2)(u_j^k - u_i^k) \end{aligned} \qquad (4.26)$$

where $\beta(i)$ is the larger neighborhood set and w is a decaying function that penalizes distance. Local anisotropic diffusion does not propagate energy across ridges, while extended diffusion does.

4.3.1.3 Steering kernels: data-adapted kernel regression

The *anisotropic* filter looks for the direction at *x* in which the image less varies, that is, the direction in which the pixel intensities are the most similar to the current one. This direction is tangent to the level line passing through x and is given by the orthogonal direction to the gradient.

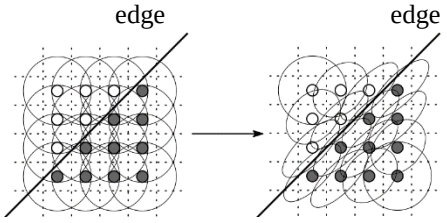


Figure 4.5: Data-adapted kernels elongate with respect to the edge.

Kernel matrix: $\mathbf{H} = \gamma \mathbf{U}_\theta \mathbf{\Lambda}_\rho \mathbf{U}_\theta^T$,

$$\mathbf{U}_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}; \ \mathbf{\Lambda}_\rho = \begin{bmatrix} \rho & 0 \\ 0 & \rho^{-1} \end{bmatrix}$$

4.3.1.4 Spatial-tonal kernels

Note that *g* and *w* may be respectively interpreted as "*certainty*" and "*applicability*" weights in the context of *Normalized Convolution* [99]. Assuming independency between domains, the usage of separate tonal and spatial kernels results in the *spatial-tonal normalized convolution* (eq. 4.27) of image pair *[v, u]* using the kernel pair *[g_s, w]*, and the variants below

$$u_i^{k+1} = \frac{\sum_{j\in\beta(i)} g_S(|v-u_i^k|^2)\,w(|x_j-x_i|^2)\,u_j^k}{\sum_{j\in\beta(i)} g_S(|v-u_i^k|^2)\,w(|x_j-x_i|^2)} \qquad (4.27)$$

- $u^{k+1} = \left[v\ , u^k\right] ** \left[K^s_{H_s}, K^t_{H_t}\right]$ *Spatio-Tonal convolution (eq. 4.27)*
- $u^{k+1} = \left[u^k, u^k\right] ** \left[K^s_{H_s}, K^t_{H_t}\right]$ *Bilateral filter*
- $u^{k+1} = \left[u^k, u^k\right] ** \left[K^s_{H_s}, 1\right]$ *Classical spatial convolution*
- $u^{k+1} = \left[u^k, u^k\right] ** \left[1, K^t_{H_t}\right]$ *Histogram transformation (converges to mean, median or mode, depending on the kernel)*

4.3.2 Stochastic Regularization

The most classical and frequent approach in image reconstruction in the image domain is the Minimum Mean Square Error (MMSE) estimation of the noise-free image which, in the Bayesian framework, is simply the mean of the posterior distribution, assuming MSE as risk [91].

4.3.2.1 Local Linear MMSE estimate in spatial domain

Almost 30 years ago, Lee [101] suggested a statistically optimal correction resulting in what could be regarded as the simplest adaptive method for image denoising. It is a two-step procedure, in which mean and variance of both noise and uncorrupted image are first estimated from a neighborhood of observed pixels, after which the pixels in the neighborhood are denoised using a standard linear MMSE. This introduced the idea that variance is a local property that should be estimated adaptively, as compared with the classical Gaussian model in which one assumes a fixed global variance [75].[86]

$$\hat{\mathbf{u}}_{LLMMSE} = E[\mathbf{u}|\mathbf{v}] = E[\mathbf{u}] + \frac{Var(\mathbf{u})}{Var(\mathbf{u}) + \sigma_{\mathbf{N}}^2}(\mathbf{v} - E[\mathbf{u}]) \qquad (4.28)$$

The adaptive behavior is controlled by the local variance. In the presence of a sharp discontinuity, the sample variance increases, decreasing the weight *w(i,j)* and causing the estimate to move toward the measured value of the pixel; that is,

[86] The approach could be regarded as *empirical* Bayes, since first the local variance is estimated and secondly this estimation is used to apply the linear (Wiener) solution locally, being optimal only when the signal and the noisy signal are jointly Gaussian and uncorrelated. This has been successfully applied in the image domain [101] as well as in the wavelet domain [110]. A formal mathematical derivation and assumptions behind can be found at [75].

less smoothing is performed. The window size is usually small, and thus the method is sensitive enough to local variations. Observe that this can also be viewed as a two-level (low-pass and high-pass) sub-band decomposition, performing a rescaling, or *shrinkage,* of the high-pass (detail) coefficients by a space varying factor depending on the likelihood (as given by the SNR) of the coefficients themselves. This provides intuitive connection to Bayesian denoising methods in wavelet domain presented in [110].

4.3.2.2 BLS formulation without explicit prior

Despite its appealing, the Bayesian approach is often criticized for reliance on knowledge of the prior distribution $p_u(u)$. If it is not known in advance, it must be learned from the uncorrupted samples (if available), or from the noise-corrupted data **v**. But, *how can a denoiser learn to denoise without having ever seen clean data?* Raphan and Simocelli show in [112] that, under restricted conditions, the Bayesian Least Squares (BLS) estimate may be written without explicit reference to the prior distribution.

Specifically, under the assumption of additive Gaussian noise,

$$p_{\mathbf{v}|\mathbf{u}}(\mathbf{v}\,|\,\mathbf{u}) = \frac{1}{(2\pi)^{n/2}\,|\,\mathbf{C}\,|^{1/2}}\, e^{-1/2(\mathbf{v}-\mathbf{u})^T\,\mathbf{C}(\mathbf{v}-\mathbf{u})} \qquad \textbf{(4.30)}$$

$$\nabla_v\, p_{\mathbf{v}|\mathbf{u}}(\mathbf{v}\,|\,\mathbf{u}) = \mathbf{C}^{-1}\, p_{\mathbf{v}|\mathbf{u}}(\mathbf{v}\,|\,\mathbf{u})(\mathbf{v}-\mathbf{u}) \qquad \textbf{(4.31)}$$

where C is the noise covariance matrix.

$$\begin{aligned}
\hat{u} &= E[U\,|\,V=v] = \int p_{u|v}(u\,|\,v)u\,du \\
&= v + \int p_{u|v}(u\,|\,v)(u-v)du \\
&= v + \frac{\int p_u(u)\,p_{v|u}(v\,|\,u)(u-v)du}{p_v(v)} \\
&= v + \frac{\mathbf{C}\int p_u(u)\nabla_v p_{v|u}(v\,|\,u)du}{p_v(v)} \\
&= v + \frac{\mathbf{C}\nabla_v p_v(v)}{p_v(v)}
\end{aligned}$$

Eq. (4.29): Bayesian Least Squares prior-free formulation.

In order to estimate $\mathbf{C}\nabla_v p_v(v)/p_v(v)$ from the noisy data v, we will use the popular nonparametric ***kernel density estimation,*** also known as the ***Parzen window(s)*** technique. Given v_i, $i=1,2,\ldots,n$ a set of n data points in the d-dimensional space R^d, the multivariate kernel density estimator with kernel $K_H(v)$ and a symmetric positive definite dxd bandwidth matrix H, computed at the point v is given by [87]

$$\hat{p}_v(v) = \frac{1}{n\,|\,2\pi H\,|^{1/2}} \sum_{i=1}^{n} K_H(v - v_i) \qquad \textbf{(4.32)}$$

The density gradient estimator is then obtained as the gradient of the density estimation by exploiting the linearity of eq. 4.32

$$\nabla_v \hat{p}_v(v) = \frac{H^{-1}}{n\,|\,2\pi H\,|^{1/2}} \sum_{i=1}^{n} (v_i - v) e^{-1/2\|v-v_i\|_H^2} \qquad \textbf{(4.33)}$$

[87] The Gaussian kernel function is perhaps the best known differentiable multivariate kernel function satisfying the conditions for asymptotic unbiasedness, consistency, and uniform consistency of the density gradient estimate [90].

$$\frac{\nabla_v \hat{p}_v(v)}{\hat{p}_v(v)} = H^{-1} \frac{\sum_{i=1}^{n} (v_i - v) e^{-1/2\|v-v_i\|_H^2}}{\sum_{i=1}^{n} e^{-1/2\|v-v_i\|_H^2}} \quad \textbf{(4.34)}$$

and, finally

$$\hat{u} = v + \frac{C}{H} m(v) \quad \textbf{(4.35)} \qquad m(v) = \frac{\sum_{i=1}^{n} v_i e^{-1/2\|v-v_i\|_H^2}}{\sum_{i=1}^{n} e^{-1/2\|v-v_i\|_H^2}} - v \quad \textbf{(4.36)}$$

Observe that m(v), known as the ***mean shift vector***, is an estimator of the normalized gradient of the underlying density and always points toward the direction of maximum increase in the density [90].

The repetitive computation of eq. 4.36 followed by the translation of the kernel according to the mean shift vector defines a procedure which leads to a local mode of the density [89][90][88]. By choosing H=C (kernel bandwidth matrix = noise covariance matrix),

$$\hat{u} = \frac{\sum_{i=1}^{n} g(|v - v_i|_H^2) v_i}{\sum_{i=1}^{n} g(|v - v_i|_H^2)} \quad g(|v - v_i|_H^2) = -k'(|v - v_i|_H^2) = e^{-1/2\|v-v_i\|_h^2} \quad \textbf{(4.37)}$$

From robust statistics (M-estimators), recall that $g(x^2) = \rho'(x^2) \Leftrightarrow \rho(x^2) = c - k(x^2)$. The solution given by eq. 4.35 corresponds to the first iteration u^0 of a gradient descent/ascent procedure to solve the following equivalent optimization problems

$$\hat{u} = \arg\min_u \sum \rho(|u - v_i|^2) \quad \textbf{(4.38)} \qquad \hat{u} = \arg\max_u \sum k(\|u - v_i\|_H^2) \quad \textbf{(4.39)}$$

which can also be interpreted as

$$\hat{u} = \arg\max_u \sum \hat{p}_v(v_i) \quad \textbf{(4.40)}$$

Observe that for these to be equivalent, $\rho(x^2) = c - k(x^2)$ and k must be the convex profile of a d-variate kernel *K(x)*. Note also that eq. 4.33 provides the value of *H* given *C* (the choice of *H* is common source of discussion. Again, the optimal bandwidth associated with the kernel density estimator is defined as the one that achieves the best compromise between bias and variance of the estimator, i.e., minimizes AMISE.

[88] *Mean shift* is a non-parametric feature-space analysis technique for locating the maxima of a density function, a so-called *mode-seeking* algorithm [89]. It moves each point in a feature space (e.g., image intensities) to a weighted average of other points using a weighting scheme that is similar to kernel density estimation, converging to the nearest mode at steady state when iterated (assuming appropriate windowing strategies). Since the algorithm does not account for neighborhood structure in images, it resembles a kind of data-driven thresholding process, particularly in the algorithm proposed by Comaniciu and Meer in [89] for image segmentation, in which the density estimate is static as the algorithm iterates.

4.3.3 Robust Regularization

In either case that the noise is not normally distributed (violation of the statistical assumption), or the local neighborhood $\beta_r(i)$ contains values from more than one distribution (violation of the continuity assumption) (i.e. ($\beta_r(i)$ contains an edge), the LS estimate is known not to be an optimal estimate due to the presence of ***outliers***. It is clear that a single of such outliers, if located sufficiently far away from the mean (which is the case of edges) can completely spoil a LS analysis. Of common practice is "*first reject all outliers according to some rejection rule, then use LS for remaining data*" as a new estimation procedure to prevent influence of distant gross errors. However, even if "good" rejection rules can be found, this clearly yields a suboptimal solution because information carried by outliers is completely discarded.

Following the spirit of previous section, an alternative to explicit modeling the prior information about the unobserved image, as is done in Bayesian regularization, is the replacement of $\boldsymbol{E_s}$ and $\boldsymbol{E_d}$ by robust error norms $\boldsymbol{\rho_s}$ and $\boldsymbol{\rho_d}$ in order to down-weight the influence of boundaries, now regarded as *outliers* of the local intensity distribution, on the estimate. This results in *robust regularization*.

This section shows one way to apply the theory of robust statistics to the data smoothing problem, which allows us to put empirical results into a wider theoretical context.

4.3.3.1 Robust Statistics

A better solution than rejection rules is the use of (classical) robust estimation techniques, which are insensitive to small departures from the idealized assumptions.[89] Classes of such techniques include *M*, *L* and *R* estimates, which correspond, respectively, to *maximum likelihood type* estimates, linear combination of *order statistics* (i.e., sorting followed by linear calculus, such as the popular weighted median filters [78]), and estimates derived from statistical *rank tests* (i.e., linear calculus followed by sorting).

For their relationship with the nonlinear filters presented in Section 4.3.1, the emphasis is here on the first type, the ***M-estimates***. Besides, they are the most flexible ones and easy to generalize, although not automatically scale invariant (i.e., they have to be supplemented for practical applications by an auxiliary estimate of scale), an issue that will be addressed in Section 4.5.

[89] The field of robust statistics [93] is concerned with estimation problems in which the data contains gross errors or ***outliers*** (i.e. "data that do not fit the pattern set by the majority of the data" [92]).

4.3.3.2 M-smoothers

Introduction of M-smoothers requires some background in the field of robust M-estimation. There is no space here to go into details. Hampel et al. [92] and Huber [93] are the classical readings for discussion and further references.

Broadly speaking, any estimate defined by a minimum problem of the form (4.41) where $\rho(x_i;\vartheta)$ is a symmetric, positive definite, arbitrary function with unique minimum, is called an M-estimate (or *maximum likelihood (ML) type* estimate[90]).

$$\min_{\vartheta}\sum_i \rho(x_i;\vartheta) \qquad \textbf{(4.41)} \qquad\qquad \min_{\vartheta}\sum_i \rho\left(\frac{x_i-\vartheta}{\sigma}\right) \qquad \textbf{(4.42)}$$

Most relevant in image smoothing is the problem of estimating a ***location*** parameter ϑ (i.e., pixel intensity), where $\varepsilon = x_i - \vartheta$ is the ***residual*** error, σ is a ***scale*** parameter that reflects the dispersion of the data set, and now ρ is referred to as ***error norm***.

A simple formulation of this problem is to consider using data $X_1,\ldots,X_n$ that are independently and identically distributed from a distribution $f(x-\vartheta)$ (where f is symmetric about the origin) to try to estimate the parameter theta. Given a function ρ, the M estimate is given by the value of theta that minimizes eq.(4.42). Applying steepest (gradient) descent to eq.(4.42) as in eq.(4.43), yields the following iterative formulation

$$\vartheta^{k+1} = \frac{\sum_i g_\sigma(x_i-\vartheta^k)x_i}{\sum_i g_\sigma(x_i-\vartheta^k)} \qquad \textbf{(4.44)}$$

This defines an estimator based on a weighted average of the data with weights depending on the sample. It is therefore a ***W-estimator*** (see [92]) and represents one possibility to obtain a solution to the local M-estimation problem.[91] Stopping after the first iteration defines a so-called ***one-step*** W-estimate or ***w-estimate***, which has been shown to be particularly efficient.

$$\begin{aligned}\vartheta^{k+1} &= \vartheta^k - \tau\frac{\partial E(\vartheta)}{\partial\vartheta}\\ &= \vartheta^k - \tau\sum_i \rho_\sigma(x_i-\vartheta^k)'\\ &= \vartheta^k + \tau\sum_i \psi_\sigma(x_i-\vartheta^k)\\ &= \vartheta^k + \tau\sum_i g_\sigma(x_i-\vartheta^k)(x_i-\vartheta^k)\end{aligned}$$

Eq.(4.43): Steepest descent applied to eq. (4.42), where $g_\sigma(\varepsilon)\varepsilon := \psi_\sigma(\varepsilon) := \rho_\sigma(\varepsilon)'$; $\rho_\sigma(\varepsilon) := \rho(\varepsilon/\sigma)$; and τ, the step size, has been chosen so that it normalizes the sum of the weights to 1.

[90] Note that the choice $\rho(x;\vartheta) = -\log f(x;\vartheta)$ gives the ordinary ML estimate.

[91] It was shown that M–estimators and W-estimators are essentially equivalent and solve the same energy minimization problem (4.42) (Hampel et al. [92], p.116). Observe however that, starting the iteration in x_i, the W-estimator converges to the local minima next to x_i if ψ is not convex. This issue is particularly relevant in multivariate and regression problems. Thus, some care is needed to ensure that good starting points are chosen, such as the *median* as an estimate of location and the *median absolute deviation* (MAD) as a univariate estimation of scale. However, for low noise cases, one commonly uses the observed data as an estimate of location.

4.3.3.3 The Influence Function ψ

To analyze the behavior of a given ρ-function against outliers, one considers its derivate, ψ, which is proportional to the so-called ***influence function*** that characterizes the bias that a sample has on the estimate.

	l_2	l_1	Gaussian	Tukey's biweight	Lorentzian (Perona-Malik)	Hubert's minimax
Error norm $\rho(\varepsilon_r)$	$\varepsilon_r^2/2$	$\lvert\varepsilon_r\rvert$	$1 - e^{-\varepsilon_r^2/2}$	$\varepsilon_r^2 - \varepsilon_r^4 + \varepsilon_r^6/3$, $\lvert\varepsilon_r\rvert \le 1$ $1/3$, $\lvert\varepsilon_r\rvert > 1$	$\sigma^2 \log[1 + \varepsilon_r^2/2]$	$\sigma(1+\varepsilon_r^2)/2$, $\lvert\varepsilon_r\rvert \le 1$ $\lvert\varepsilon_r\rvert$, $\lvert\varepsilon_r\rvert > 1$
"Influence function" $\psi(\varepsilon_r)$	ε_r	$sign(\varepsilon_r)$	$\varepsilon_r/\sigma \cdot e^{-\varepsilon_r^2/2}$	$\varepsilon_r [1 - \varepsilon_r^2]^2$, $\lvert\varepsilon_r\rvert \le 1$ 0, $\lvert\varepsilon_r\rvert > 1$	$\varepsilon_r / (1 + \varepsilon_r^2/2)$	ε_r, $\lvert\varepsilon_r\rvert \le 1$ $sign(\varepsilon_r)$, $\lvert\varepsilon_r\rvert > 1$
Weighting function $g(\varepsilon_r)$	1	$1/\lvert\varepsilon_r\rvert$	$e^{-\varepsilon_r^2/2}$	$\frac{1}{2} [1 - \varepsilon_r^2]^2$, $\lvert\varepsilon_r\rvert \le 1$ 0, $\lvert\varepsilon_r\rvert > 1$	$1/(1 + \varepsilon_r^2/2)$	$1/\sigma$, $\lvert\varepsilon_r\rvert \le 1$ $1/\lvert\varepsilon_r\rvert$, $\lvert\varepsilon_r\rvert > 1$
Scale	*(indep. of scale)*		$\varepsilon_r' = \varepsilon_r$	$\varepsilon_r' = \varepsilon_r/\sqrt{5}$	$\varepsilon_r' = \sqrt{2}\,\varepsilon_r$	$\varepsilon_r' = \varepsilon_r$
Result	*mean*	*Median*	*mode approximation*			

Table 4.1: Error norm comparison. For convenience, we use the *relative* residual error notation $\varepsilon_r := \varepsilon/\sigma$. Values of ε_r' chosen as a function of ε_r so that outlier "rejection" begins at the same value $\varepsilon_r = 1$ for each function.

In the location case, monotonicity and boundedness of ψ ensure, respectively, that *(i)* the set of solutions of the implicit equation is unique (or at least convex), and *(ii)* outliers have bounded (though maximum) influence [92]. To further reduce the influence of outliers with respect to the other samples, one can use smoothly *redescending* M-estimators, such as Tukey's biweight, with ρ being bounded and ψ continuously becoming zero for large $|\varepsilon|$. They completely reject distant outliers, but not suddenly, allowing a transitional zone of increasing doubt, and are therefore much more efficient than "hard" rejection rules. However, because the influence function is no longer monotonic, they may lead to multiple solutions

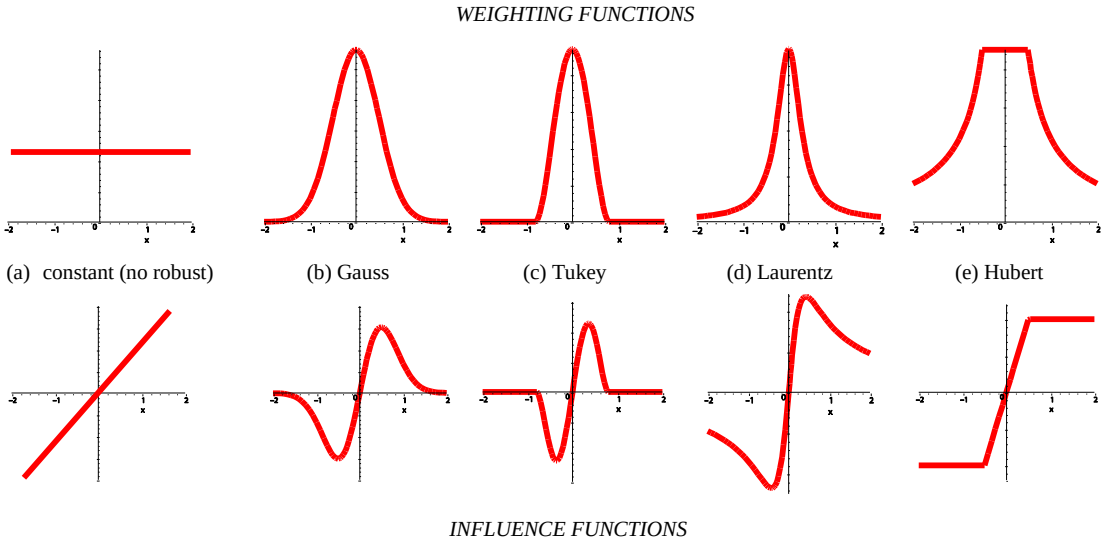


Figure 4.6: Sample weighting functions and their corresponding influence functions, scaled so that outlier "rejection" begins at the same value for each function. Reproduced from [86].

It can be shown that the WLS is a particular form of the M-estimator and, that in both, the weight function can be regarded as the membership function in fuzzy set theory. The robust estimate, therefore, represents the cluster center (prototype) and the membership value (weight) of a point is determined by its distance from the prototype. During the minimization process, a sample that is far from the prototype, i.e., an outlier, will be treated less importantly and vice versa. From this point of view, the use of a weighted, or equivalently fuzzy or probabilistic, LS method for the design of the image smoothing filter yields robust results.

It is interesting to note that common robust error norms have frequently been proposed in the literature without mentioning the motivation from robust statistics. For instance, edge-stopping functions of anisotropic diffusion presented in section 4.3.1 serve the same role as robust energy functions, the use of explicit edges in Bayesian regularization is equivalent to a robustification of the prior by robust error norms, and global M-estimators (a.k.a. *histogram operators*), where no location data is used, can be considered either as a particular case or a step towards mean-shift. Also, all these methods have been casted into a unified framework for functional minimization combining nonlocal data and nonlocal smoothness terms in [104], as shown below.

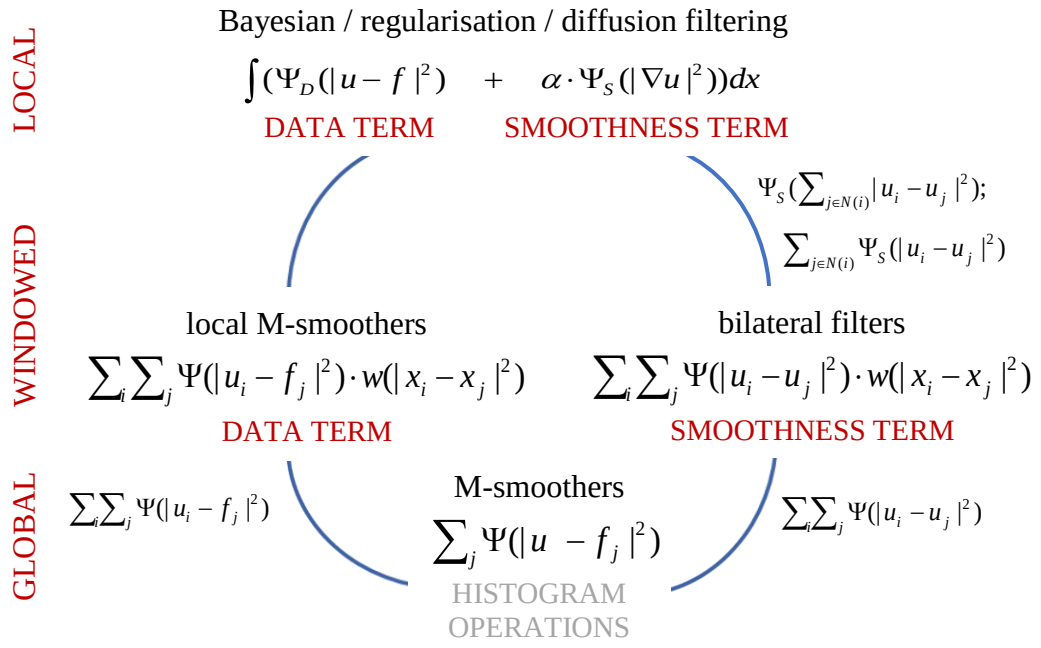


Figure 4.7: Overview of the methods studied in this chapter and the energy functional that they respectively minimize. Starting from *regularization methods* fitting into the *Bayesian framework* at the top, we went clockwise down to the right branch concentrating on the smoothness term only. We estimated the gradient magnitude |Vu| using discrete samples, extended the size of the estimation window, resulting in *anisotropic diffusion*, and derived the *bilateral filter* on the right. By extending the spatial window size, the circle can be closed to *statistical M-estimation* and *histogram-based* global methods, which can in turn be made local by introducing a spatial weighting window *w*. Adapted from [104].

4.4 State of the art: *Neighbourhood* Filters

The term *neighbourhood filters* refers to all image filters which reduce the noise at pixel i by averaging pixels based on their vicinity to i, as measured by $g_v(i,j)$. Here the notion of *neighbourhood* must be understood broadly: nearby spatial locations, as in *domain filtering*, which enforces closeness by weighing pixel values with coefficients that fall off with distance, regardless of their actual value (and thus blurring edges); similarity in the photometric range, as in *range filtering*, which averages image values with weights that decay with dissimilarity to the centre pixel, regardless of their spatial position; or feature closeness, as in *patch-based filtering*, which averages pixels having a similar neighbourhood.

4.4.1 Range filtering

Under the fairly general assumption of a generalized signal-dependent additive noise model,[92] denoising can be achieved by first finding out the pixels $J(i)$, which received the same original energy (i.e, $j \in J(i) \Leftrightarrow u(j) = u(i)$) and then averaging their observed grey level [83] (depending on the noise model, other statistical estimates are of course possible like the median, etc.). Here the challenge is finding $J(i)$ for every i, since the original image value $u(i)$ is lost. While the simplest idea would be to assume that all pixels with the same observed value $v(i)$ have the same noise model, this would result in no filtering effect. Instead, range filters first define a measure of similarity between different regions in the image, and then compute the denoised value $NFu(i)$ at pixel i as a weighted average of all pixels in the image, $v(j)$,

$$NFu(i) := \frac{1}{|J(i)|}\sum_{j\in J(i)} v(i) \qquad NFu(i) = \frac{1}{C}\sum_{j\in\Omega} g_v(i,j)\cdot v(j); \qquad C = \sum_{j\in\Omega} g_v(i,j)$$

where the weights $g_v(i,j)$ depend on the similarity between pixels i and j, and C acts as a normalization to ensure that a constant function is mapped to itself.[93]

Typically, the similarity is expressed by a dissimilarity measure $d_v^2(i,j)$, and $g_v(i,j)$ is a decreasing function with scale parameter h, which controls how fast the weight decay with increasing dissimilarity, thus acting as a degree of filtering:

$$g_v(i,j) = \exp\left(-\frac{d_v^2(i,j)}{h^2}\right) \quad \text{and} \qquad d_v^2(i,j) = \| v(j) - v(i) \|^2$$

Range filters are nonlinear because their weights depend on image intensity or colour. Computationally, they are no more complex than standard non-separable filters. Most importantly, they preserve edges. Since similar pixel values can be located far from each other, this leads to an essentially nonlocal filtering. Such have no notion of space, they merely transform the image's intensity histogram.

---

[92] *Remark:* according to generalized signal-dependent additive noise model (briefly presented in section 4.1.2), the noise $n(i)$ at each pixel i only depends on the original pixel value $u(i)$ and is additive, i.i.d. for all pixels $j \in$ J(i) with the same original value as i.

[93] Observe that, by the normalization C, only the relative difference in similarity is considered.

4.4.2 Local Neighbourhood filters: the Bilateral Filter

Given now an image where all the pixel values are equally likely, i.e., with a flat intensity histogram (e.g., resulting from applying *histogram equalization*, a popular image enhancement technique), the filtered image would be the same as the noisy one, since the dissimilarity function $d_v^2(i,j)$ is symmetric. Put formally, the *a posteriori* probability $p_v(v)$ is non-informative. In other words, pixel values are not enough for computing similarity between pixels. Instead, feature (e.g., neighbourhood) histogram is much sparser, thus much more informative, and the resulting filter can be made non-local and still be robust.

The fact that spatial locality is still an essential notion leads to *local* neighbourhood filters, a special case of data-adaptive kernel regression, based on a generalization of the Nadaraya-Watson estimator[94] with the spatio-tonal or bilateral kernel, $K_{Hs}(x_i - x) \cdot K_{hr}(y_i - y)$, whose role will be to enforce both photometric and geometric locality. Such a combination of domain and range filtering, from now on referred to as ***bilateral filtering***, has proven to be a powerful tool for adaptive denoising purposes, due to its good edge-preserving properties [86][107][116].

The bilateral filter may be regarded as the first modern adaptive method to successfully suppress noise without loss of finer details. It can be attributed to Tomasi et al., where the authors proposed a generalization of the SUSAN filter, which itself was an extension of the Yaroslavky filter [116].

As shown for one image scan-line in the figure 4.9, we can approximate the extent of the combined spatial and range filters as a rectangle (or better an ellipse) centred around each input pixel.

The rationale of bilateral filtering is that two pixels are close to each other not only if they occupy nearby spatial locations but also if they have some similarity in the photometric range. Note that use of separate tonal and spatial kernels assumes independency between domains. A more general approach considers linear filtering in higher-dimensional space, using a single constant weight function based on the Euclidean distance defined on the joint spatial-tonal domain S×R [79].

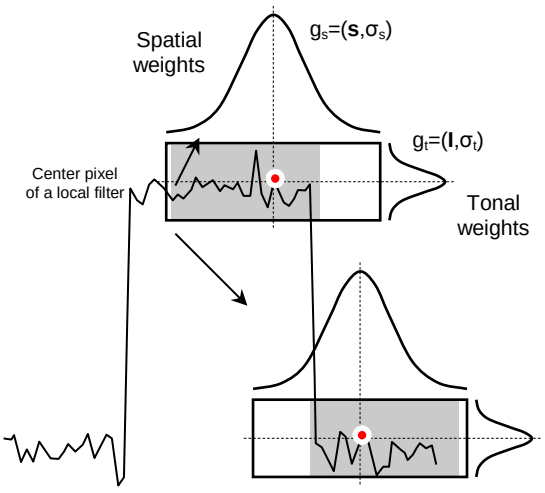


Figure 4.8 A useful way to understand the main idea is to view it as a variation of local weighted averaging. Most smoothers are essentially local weighted averages in the x direction (of a scatterplot), but the sigma filter also applies local weights in the y direction.

[94] I.e., the 0th order estimator from the classic kernel regression framework presented in section 4.2.2.

4.4.2.1 Theoretical Studies

In parallel to applications, a wealth of theoretical studies have explained and characterized the bilateral filter's behaviour, relating it to a broader class of known non-linear filters such as anisotropic diffusion and robust estimation. This new insight has served for improving the bilateral filter and extend its use for other applications.

4.4.2.1.1 Connection to PDEs

Several authors have shown in [75] and [86] that the bilateral filter restricted to the four adjacent neighbours of each pixel actually corresponds to a discrete version of Perona and Malik anisotropic diffusion model [108] (see Section 4.3.1.1). This result has been extended by Elad [87] and Barash and Comaniciu [79] who demonstrated that anisotropic diffusion solvers can be extended to larger neighbourhoods, that is, the image derivatives are computed with pixels at a distance, not only with adjacent pixels[95], thus producing a broader class of extended nonlinear diffusion filters, which includes iterated bilateral filters as one special case. Hence, while bilateral filtering is a single-pass operation, anisotropic diffusion takes many iterations to achieve the same result. Bilateral filtering also reduces noise more effectively because its filter support can reach beyond a ridge barrier. This is something anisotropic diffusion cannot do because the diffusion simply stops at high-gradient barriers. Finally, while they both respect causality (no maximum or minimum can be created, only removed) only anisotropic diffusion is adiabatic (i.e., energy-preserving).

4.4.2.1.2 Connection to Robust Statistics

In a similar manner to the work of Black et al. [81] on PDE filters, several authors have studied the bilateral filter in the framework of robust statistics [118]. They showed that the bilateral filter (and neighbourhood filters in general) is a *w-estimator* (see Section 4.3.3.1), which results from replacing the L_2 norm in the Nadaraya-Watson estimator by a robust one:

$$\hat{m}_{NW}(x_i) = \arg\min_{\vartheta} \sum_{j=1}^{n} (Y_j - \vartheta)^2 K_h(x_i - x_j) \;\rightarrow\; \hat{m}(x_i) = \arg\min_{\vartheta} \sum_{j=1}^{n} \rho(Y_j - \vartheta) K_h(x_i - x_j)$$

This explains the role of the tonal weight in terms of sensitivity to outliers.

Neighbourhood filters are also referred to as local M-smoothers when iterated [104]. However, there is one significant difference: local M-smoothing uses the initial image in the averaging procedure and searches for the steady state, while bilateral filtering uses the evolving image and has to stop after a certain number of iterations in order to avoid obtaining a flat image.

---

[95] In a finite-difference implementation of anisotropic diffusion, the gradient magnitude is computed from intensity differences between the centre pixel and its direct neighbours. The edge-stopping function is then analogous to the tonal weight in bilateral filtering, both prevent samples from distant intensity modes to participate in the local average. The spatial weight of the bilateral filter, on the other hand, covers the same Gaussian support as in isotropic diffusion.

4.4.2.1.3 Connection to Local Histograms, mode seeking and Mean Shift

Van de Weijer and Van den Boomgaard [117] demonstrated that the result of the bilateral filter at a given pixel is the average intensity of its local histogram with each bin weighted by the range function.[96] This selective sampling of the *LOIs* along the tonal axis is mode-seeking in nature, i.e., when iterated, it replaces the intensity of a pixel by its closest local mode. The edge-preserving capability is due to an effective selection of the closest mode in the local histogram as defined by the filter support.[97] This behaviour can be interpreted in terms of robust statistics: pixels in the same mode are considered as inliers whereas pixels in other modes are outliers, i.e. ignored. Finding the mode in the histogram that is most likely to represent the distribution that the point belongs to, leads to ***local mode filtering***, a technique already proposed in the early eighties that is shown to result in visually impressive results.

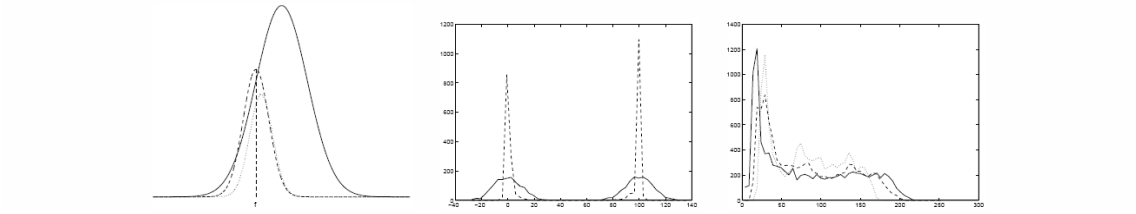


Figure 4.10 The bilateral filter is driven by the modes of the local histograms. Remember that adding noise is like convolving the image histogram with the noise pdf, what smoothes it. This process can be reversed by moving each point upwards to its local mode.

4.4.2.1.4 Connection to Linear Filtering in higher-dimensional spaces

Sochen et al. [113] have introduced the notion of image manifolds where an image I is represented by a manifold M embedded in the joint spatial-range domain $S \times R$:

$$(p_x, p_y) \in S \rightarrow M(p_x, p_y) = (p_x, p_y, I(p_x, p_y)) \in S \times R$$

In this context, the bilateral filter is shown to be related to the short-time kernel of the heat equation defined directly on the image manifold. Barash showed in [79] that the two weight functions are actually equivalent to a single weight function based on the Euclidean distance defined on $S \times R$, instead of the manifold geodesic distance. Based on this geometric interpretation, he related the bilateral filter to adaptive smoothing. Using a similar approach, but in a signal-processing context, Paris and Durand have demonstrated in [107] that the bilateral filter corresponds to a Gaussian convolution (i.e. linear filtering) in this higher-dimensional, homogeneous space.

---

[96] *Local histograms* are classical intensity histograms where each pixel contributes only a fraction defined by a spatial influence function. The set of such local histograms over the whole image is called *Local Orderless Images* (LOIs) [98].

[97] While multiple iterations of bilateral filtering might be required before the process converges to the true local mode, the first iteration of bilateral filtering already covers a significant step towards the local mode that further iterations may not be necessary.

4.4.2.2 Extensions

The strengths and limitations of bilateral filtering are now fairly well understood. Therefore, several extensions have been proposed [83][84][87]. Two main directions have been followed: first, variants have been developed to better handle gradients by taking the slope of the signal into account; second, bilateral filtering has been extended to handle several images in order to better control the way edges are detected.

4.4.2.2.1 Accounting for the local slope: high-order Bilateral Filters

Most authors noted that the zeroth order bilateral filter implicitly assumes that the desired output should be piecewise constant, and thus is particularly good at preserving step-like edges.

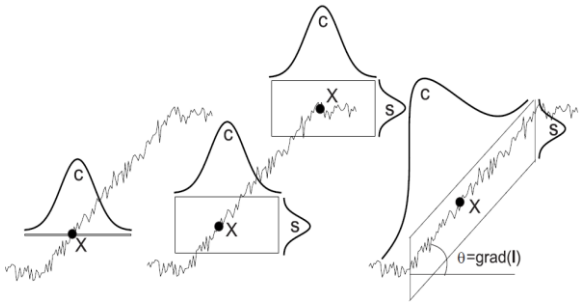


Figure 4.11: Filter extent for one scanline of an image. Reproduced from [84].

In order to improve edge-preserving results for ridge- and valley-like edges as well, several studies have proposed extensions to the bilateral filter, smoothing towards piecewise constant-gradient (or low curvature) results instead [83][84][107]. E.g., Choudhury and Tumblin [84] "tilt" the filter extent of a bilateral filter applied to image intensity; this affine transform of the range filter, a.k.a *trilateral filter*, as shown in Figure 4.11, restores the effectiveness of the spatial filter term.

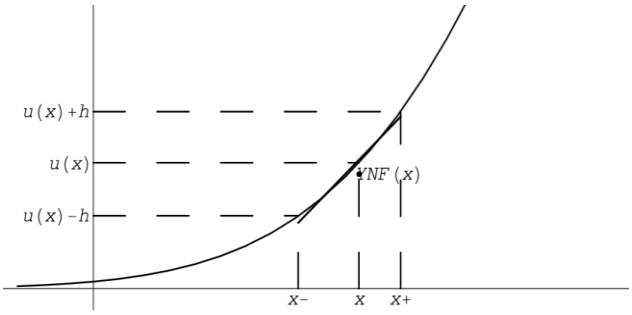


Figure 4.12: Neighbourhood filters create stepwise functions. The reason for the staircase effect is that for each x, the number of points y such that $u(x)-h < u(y) \leq u(x)$ is larger than the number of points satisfying $u(x) \leq u(y) \leq u(x)+h$. The regression line of u inside $(x-, x+)$ better approximates the signal at x.

4.4.2.2.2 Using several images: Cross and Joint Bilateral Filter

In computational photography applications, it is often useful to decouple the notion of edges to preserve from the image to smooth. With this purpose in mind, Eisemann and Durand introduced the ***cross bilateral filter***, also known as the ***joint bilateral filter,*** as a variant of the bilateral filter [107].

Given an image $\boldsymbol{I}$, the cross bilateral filter smoothes it while preserving the edges of a *guidance* image $\boldsymbol{E}$, which is used to compute the tonal weights. $\boldsymbol{E}$ is typically less noisy, and thus more reliable than $\boldsymbol{I}$ (e.g., the same scene taken with flash light).

4.4.2.2.3 Extension to colour images

Neighbourhood filters can be applied to colour images just as easily as they are applied to black-and-white ones. The CIE-Lab colour space, already introduced in Chapter 3, endows the space of colours with a perceptually meaningful measure of colour similarity, in which short Euclidean distances correlate strongly with human colour discrimination performance [78]. Thus, if we use this metric in our bilateral filter, images are smoothed and edges are preserved in a way that is tuned to human performance. Only perceptually similar colours are averaged together, and only perceptually visible edges are preserved.

4.4.2.3 Disadvantages of neighbour filters

The problem with these filters is that comparing only grey level values in a single pixel is not so robust when the standard deviation of the noise exceeds the smallest feature contrast. Neighbourhood filters also create artificial shocks which can be justified by the computation of its method noise, as noticed in [83]. Moreover, they assume that a given centre pixel is a prototype of its neighbouring pixels. Therefore, if the given centre pixel itself is a noise pixel, the assumption is not valid and consequently the filter will not work well. In [115], Takeda et al. proposed a signal-dependent steering kernel regression (SKR) framework for denoising, which proved to be much more robust under strong noise.

In general, denoising based on the similarity between single v values may be insufficient, especially with images containing many structured patterns, which can be misclassified either as details to be preserved or noise, leading to artefacts and blurring effects. Similarity is much more reliable if it is evaluated by comparing a whole window around each pixel, not just the colour of the pixel itself, what significantly reduces the misidentification probability.

Remark also that breaking into spatial and radiometric terms as utilized in the bilateral case weakens the estimator performance since it limits the degrees of freedom and ignores correlations between positions of the pixels and their values. Fortunately, bilateral filtering provides a direct measure of this uncertainty: the normalization factor k is the sum of the influence of each pixel, which can therefore be used to detect dubious pixels that need to be fixed. In [86], Durand and Dorsey propose to use the log of this value because it better extracts uncertain pixels.

Finally, while local neighbourhood filters are derived from local regularization, the most similar pixels to a given pixel have no reason to be close to it (think, e.g., of periodic patterns, or of the elongated edges which appear in most images). This observation has led to the development of non-local neighbourhood filters discussed in the next section.

4.4.3 Non-local Neighborhood filters: Non-Local means

Recently, the translation self-similarity of images has been exploited for the purpose of denoising, replacing conventional local neighbourhoods with data-driven non-local estimation domains, where the mutual similarity between different local regions determines the weights or the shape of the non-local domain [83]. The fact that such a self-similarity exists is a regularity assumption, actually more general and more accurate than all regularity assumptions we have already considered.[98]

Inspired by previous exemplar-based approach for texture synthesis, Buades et al. [83] have recently extended neighbourhood filters to a wider class, the *generalized neighbourhood filter*, which they called *non-local means* (NL-means)

$$\hat{u}(x_i) = \frac{\sum_{x_j \in \Omega} L_g(\mathbf{V}_i - \mathbf{V}_j) K_h(x_i - x_j) v_j}{\sum_{x_j \in \Omega} L_g(\mathbf{V}_i - \mathbf{V}_j) K_h(x_i - x_j)} \qquad \textbf{(4.45)}$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$ and $L_g(\cdot) = (1/g)L(\cdot/g)$ are rescaled versions of non-negative kernel functions and $\mathbf{V}_j$ denotes a vector of pixel values taken in the neighbourhood of a point x_i belonging to the image domain.

The similarity between two points x_i and x_j is measured not by their single v-value but rather by the Euclidean distance $||\mathbf{V}_i - \mathbf{V}_j||^2$ between two vectorized image *patches*

$$d_v^2(i,j) = \| v(j) - v(i) \|^2 \quad \rightarrow$$
$$d_v^2(i,j) = \| v(N_j) - v(N_i) \|_{2,a}^2$$

where $\| I \|_{2,a}^2 = \sum_{\mathbf{x} \in \Omega} G_a(\mathbf{x}) |I(\mathbf{x})|^2$, and G_a is a two-dimensional Gaussian kernel of standard deviation a, centred at $(0,0)$, and of the same dimension as I –e.g., $dim(I)=3$ for colour images-). This yields a more robust comparison than previous neighbourhood filters, as illustrated in the figure, allowing to remove noise even from textured images without destroying the fine structures of the texture itself. Finally, the spatial support of the filter is controlled by h and the level of blurring by g. Both parameters are set manually according to the image contents and the signal-to-noise ratio.

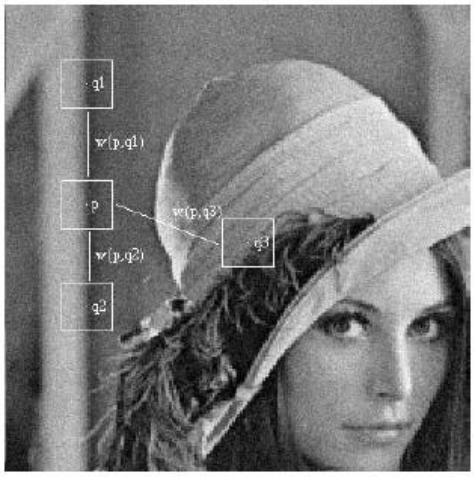


Figure 4.13: the pixel q_3 has the same grey level value of pixel p, but the neighbourhoods are much different and therefore the weight $w(p, q_3)$ is nearly zero. Reproduced from [83].

[98] In contrast to *smoothness*, which is a *local* regularity property of natural images, *self-similarity* is an example of a *nonlocal* one, in the sense that local neighbourhood of an image can be highly correlated (i.e., affinely similar) to other neighbourhoods through the image.

The NL-means filter may be regarded as a generalization of previous neighbourhood filters with the systematic usage of all possible self-predictions the image can provide. This observation allows us to bridge NL-means to diffusion, non-parametric estimation and robust statistics. Indeed, if the size of the patch is reduced to one pixel and $L_g(\cdot)$ and $K_h(\cdot)$ are Gaussian kernels, the NL-means filter is then equivalent to bilateral filtering. Besides, following the connection of neighbourhood filters and local mode seeking, the denoising effect of this algorithm can be understood in similar terms: as it forces the probability density to concentrate, groups of similar windows tend to assume a more and more similar configuration which is less noisy. Some authors have noticed that, under stationarity assumptions, for a pixel *i,* the NL-means algorithm converges to the conditional expectation of *i* once observed a neighbourhood of it.

4.4.3.1 Performance

Even though the algorithm is extremely simple, essentially described by the equation above, it has demonstrated strong superiority over local-based spatial methods such as bilateral filter in terms of both PSNR and visual quality, especially for texture-like images containing many repeated patterns, like natural images, since it is able to separate texture, edges and high frequency signals from the noise with the resulting residuals typically looking like pure noise with no or low correlation to the noisy image and showing almost no texture nor other structure.

One of the limitations of the NL-means algorithm is the removal of highly structured noise as in jpeg compressed images, where it can remove the block artefact due to compression but at the cost of removing some details as the difference between the compressed and restored images shows.

4.4.3.2 Iteration

Some authors have proposed using the nonlocal means filter in an iterative manner, where the filtering result is employed to redefine the similarity of patches in the next iteration: rather than imposing similarity of $u(x)$ to $v(y)$ for locations y where the input image $v(y)$ is similar to $v(x)$, they impose similarity to $v(y)$ for locations y where the filtered image $u(y)$ is similar to $u(x)$.

$$NL(u(i)^{k+1}) = \frac{1}{C}\sum_{j\in\Omega} g_{u^k}(i,j)\cdot v(j) \qquad (4.46)$$

This induces an additional feedback and further decouples the resulting image u from the input image v [82]. The idea is that the similarity of patches can be judged more accurately from the already denoised signal than from the noisy input image. The iteration of symmetric NL-means may be interpreted as a heat equation on the patch manifold. Experimental results demonstrate that the iterated nonlocal means filter, while prone to *hallucination* of regular patterns, outperforms both nonlocal means and total variation filtering when applied to the restoration of regular textures.

4.4.3.3 Extensions

Some authors further propose to replace the neighbourhood weighting in the original formulation by a sorting criterion, which assures that the amount of filtering no longer depends on how repetitive respective image structures are in the given image. This addresses the parameter selection problem of the original nonlocal means filter and leads favourable denoising results of textured images, particularly in case of large noise levels [82]. This is discussed in more detail in section 4.4.4.

4.4.3.4 Implementation details

In [83] it is recommended to set $h \approx 12\sigma$. This adjustment of the decaying parameter h to a value higher than the expected value $\sqrt{2}\sigma$ is probably related to the fact that the two compared patches are not independent. Note that some pixel values are in common in the two vectors but at different locations.[99]

The point on using a soft Gaussian threshold for the weights $w(x,y)$ is that we may find pixels for which there is no identical or nearly identical window in the image. In that case, the threshold strategy should leave exactly the noise value at such points. The result would visually be identified as an impulse noise and the noise to noise principle would be violated. An exponential function is used instead of the threshold and makes a more adaptive weighting distribution.
In order to involve the current pixel in its own average, the distance between the window centred at the reference pixel and itself is set equal to the minimum of the other distances. Otherwise, the probability distribution should be excessively large at the pixel itself.

Compared to other denoising algorithms which have $o(n^2)$ complexity where n^2 is the number of pixels in the image, these algorithms have $o(n^4)$ time complexity, which prevents it from being used in real applications. For computational purposes, the search of similar windows is restricted in a large *"search window"* of size S×S pixels. Usually a search window of 21×21 pixels and a similarity square neighbourhood N_i of 7×7 pixels is used. If N^2 is the number of pixels of the image, then the final complexity of the algorithm is about $49 \times 441 \times N^2$. The 7 × 7 similarity window has shown to be large enough to be robust to noise and small enough to take care of details and fine structure. Further improvement can be achieved by multiresolution strategies, as been proposed in [83].

[99] In [83], it is implicitly assumed that $v(x_i) \mid v(x_j) \sim N(v(x_j), \frac{1}{2}h^2 I_n)$. Actually, this hypothesis is valid only for non-overlapping and statistically independent patches, but most of patches overlapped.

4.4.4 Bandwidth issue in Neighbourhood filters

While neighbourhood filters, and particularly the non-local means, can yield astonishing denoising result, several empirical studies have revealed a large sensitivity to the scale parameter h, which is responsible for steering the decay of weights for decreasing similarity of patches, to separate the good data (inliers) that fit the model, from the gross errors (outliers). This parameter sensitivity increases with the noise variance σ^2 in the image. Moreover, it is often found that if the noise level exceeds a certain value, it is no longer possible to choose a global h such that the noise is removed everywhere without destroying structure somewhere else in the image.

The reason for this effect is that, while with a highly repetitive patch (and a rather small noise level), there will be many similar patches for which $g(x,y) \approx 1$, with a patch that is hardly similar to other patches in the image (or only few of them), there will be almost no change at x since $g(x,x) = 1$ and $g(x,y) \approx 0$ almost everywhere. In this case, one has to increase h such that there are enough y with $g(x,y) > \varepsilon$ in order to see a smoothing effect.

Buades et al. have been aware of this problem and suggested to set $g(x, x)$ to $max_{y \neq x}\, g(x, y)$. Although this attenuates the problem, it does not resolve it, as it only ensures the averaging of at least two values, which in many cases is not sufficient.

In robust estimation one frequently needs an initial or auxiliary estimate of scale. For this one usually takes the median absolute deviation $MAD_n = 1.4826\, med_t\{|x_t - med_j x_j|\}$, because it has a simple explicit formula, needs little computation time, and is very robust as witnessed by its bounded influence function and its 50% breakdown point.

Approaching the problem from a different point of view, [82] have proposed to choose the number n of positions that is appropriate to remove a certain noise level, replacing the neighbourhood weighting by a sorting criterion. They then simply take those n patches with the smallest dissimilarity $d^2(x,y)$. By considering for any pixel x the n most similar pixels rather than all those pixels of similarity above a fixed threshold, we allow for denoising which does not depend on how repetitive the respective structure at x is in the given image. This addresses the parameter selection problem of the original nonlocal means filter and leads to favourable denoising results of textured images, particularly in case of large noise levels.

We see that this solution is simply the ***K Nearest Neighbour*** (KNN) approach to the problem of bandwidth selection in kernel regression, as a natural and easy alternative to ***fixed*** bandwidth.

4.5 Noise Level Estimation

In many algorithms for image processing tasks such as restoration, edge detection or image segmentation, a precise estimate of the type and amount of noise present in the image is required in order to achieve an optimal result, since it allows to: ***i)*** assess reductions in image quality as a consequence of the degradation process; ***ii)*** select the appropriate IP strategy; ***iii)*** set initial values of tuning parameters; and ***iv)*** adapt to the local noise characteristics instead of using fixed ones for the whole image.[100] A perhaps unexpected application of noise level and correlation estimation is the analysis of the image processing occurring inside cameras. From the observed characteristics of the noise and a priori knowledge of its initial form, we may infer the transfer function of the device. This serves for further purposes (e.g. estimation of brightness corrections or compression algorithms applied).

Consequently, the ability to accurately estimate the noise level (in general the properties of the degradation) contaminating images is central to both image quality assessment and restoration, as well as an essential step toward achieving reliable, fully automatic computer vision algorithms. However, compared to the in-depth and wide literature on image denoising, the literature on noise estimation is rather very limited. This work looks in that direction and extends both classical [105] and very recent [103] previous work on non-supervised techniques which, given the observed corrupted image, provide us with relevant noise estimates for image quality assessment and subsequent IP steps.

4.5.1 The blind noise variance estimation problem

In practical situations, where a priori noise characterization is hardly available and it is expensive, difficult or even impossible to obtain noise-free reference images from which to estimate the noise as the difference with the noisy observations, the alternative is to estimate its statistics directly from the noisy observations. In such a case, the problem translates into measuring deviations in intensity from an ideal image that may contain structure.[101] When multiple observations are available, noise estimation is an over-constrained problem, comparable to image denoising by averaging several images. An efficient blind noise variance estimation algorithm should return the correct noise parameters for a large range of values (including the case of an uncorrupted image when we expect to obtain zero as the estimated variance of the noise), different images, and within a reasonable time, compared to the subsequent processes (e.g.,

[100] Alternatively, noise characterization allows noise equalization (normalization) and whitening (decorrelation), so that the resulting image well meets the AWGN assumptions. This approach could be referred to as implicit approach to non-AWGN denoising, since the algorithms remain unchanged, but the image is transformed to meet assumptions. The idea comes from chapter 3.7 in [78].

[101] *Structure*, *texture*, *regularity*, correlation is what allows us to tell apart natural images from noise.

denoising or segmentation). Generally, the accuracy of the proposed algorithms is not always satisfactory, especially for low noise levels.

Noise estimation from a single observation of a corrupted image is an under-constrained problem, and thus frequently regarded as a much more challenging task, requiring further assumptions to be made. However, in contrast to image denoising, noise characterization does not require finding the noise image, but just a statistical (typically 2nd order) characterization. This greatly simplifies the task, especially if one assumes, as typically occurs, that many realizations of the noise process can be found within a single image.

In image denoising literature, most estimation methods traditionally assume additive Gaussian noise. If it is further assumed to be white, the problem is relatively easy, because one just need to estimate its variance to have a complete statistical characterization. This can be done using a decorrelating linear transform (e.g., Fourier, wavelet, PCA) for decoupling signal and noise, according to their different spectral features. To this kind of methods belongs the classical approach of estimating the noise variance as the smallest eigenvalue of the sample covariance matrix or, in a more advanced version, the use of a robust statistic at the output of a high-pass wavelet sub-band [1].[102]

Most of the various algorithms operating in the image domain proposed in the literature for estimating the variance of additive noise fall into two main categories [105]:

a) those which ignore heterogeneous regions and use the remaining pixels, initially classified as showing little structure, to estimate the variance [101].
b) those which filter the noisy image first to suppress structure and then estimate the variance from the residuals [94].

While methods in the former category tend to underestimate the variance, as avoiding heterogeneous regions would naturally bias them to less noisy regions, the latter tend to overestimate the variance, as they are unable to completely remove structure. For each of the groups most methods show strong similarities. While some of them are of a pure heuristic nature, other are statistically more solidly founded.

All methods in the first category are based on the observation that in uniform regions of the image, the variations are mainly due to noise. Noise variance is then estimated by computing a local measure in these uniform regions and deriving a, possibly global, estimate from these measurements. A commonly used method for estimating noise variance is to identify large homogeneous image areas and use them to calculate the noise statistics. The problem with this method is that it must be supervised, the image might not have large

[102] Maximum Likelihood estimation of noise variance under known normalized power spectra of noise has been related in [110], where a generalized expectation maximization algorithm is proposed to estimate spectral features of a noise source corrupting an observed image.

homogeneous areas, and a large number of areas with varying means are necessary to define a linear noise model [102].

Alternatively, one may ***i)*** divide the image into small blocks; ***ii)*** measure the intensity variations for every block; and ***iii)*** assuming that the noise variance is much smaller than that of the image, the block with the least variation (or the average of the blocks with the smallest variation) should correspond to a constant brightness region. The main trade-off is, as before, in window size: to avoid even small scale image structures from contributing the noise measure, one would choose a small (e.g. 3x3 pixels) support area. However, the smaller the support area, the bigger the estimator variance.

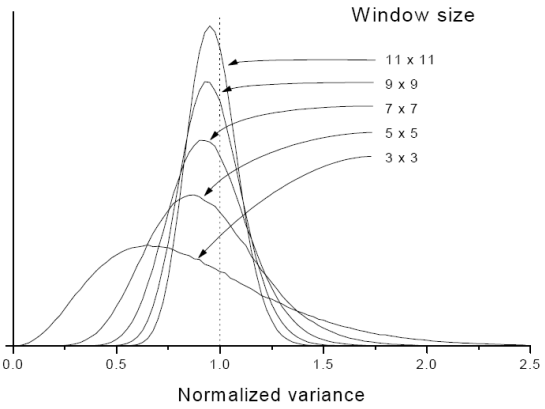


Figure 4.14 Distribution of computed local variance for different window sizes

4.5.1.1 Lee's method (1st category)

While most of these approaches assume of constant noise characteristics, only a few authors, especially Lee et al. [102] and [103], explicitly model the noise variance. Lee et al. [102] assume a multiplicative as well as an additive noise component.

First, the average m and the variance v of the intensity are computed for all image blocks of the size of 8x8 pixels. Then, assuming a large number of small blocks correspond to homogeneous areas, a scatter plot of v versus m^2 should reveal a primary cluster to which a straight line may be fitted using a Hough transform [78] (to make the estimation robust with respect to image blocks showing structure, which will have a higher variance and be sparsely scattered above the primary cluster in the scatter plot). Finally, the variances of the multiplicative and additive noise component are estimated by the slope and intercept of the fitted line. For the case of purely additive noise Lee et al. suggest using a scatter plot of the standard deviation versus mean.

The Hough transform can be avoided by estimating σ_n by the larges peak (i.e., the mode) in a histogram of the block standard deviation. Figure 4.15 shows a histogram of all local standard deviations in Figure 4.13 having values from 0 to 100. If Figure 4.13

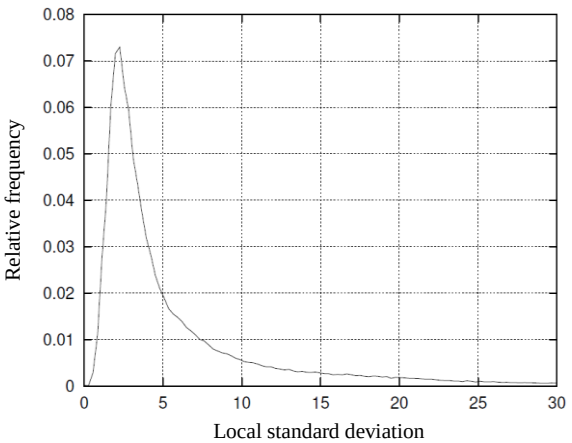


Figure 4.15: Histogram of local 3x3 standard deviations from image in Figure 4.13.

was completely homogeneous, this histogram would be normal with mean σ and variance $\sigma^2/9$.

However, the existence of 3x3 heterogeneous regions places many high variance entries in the right-hand tail of the histogram, as noticed in [111]. Now the left part of the histogram referring to small values reflects the effect of noise whereas the right part reflects the effect of edges or textured areas. However, despite these outliers, the histogram still has a clear peak at around σ, the standard deviation of the added noise. Thus, the mode of the local standard deviation distribution could be used as a *robust* estimate for σ[103].

4.5.1.1.1 Non-parametric methods for the computation of the mode

a. ***Histogram-based methods***

The most common mode estimation method for discrete or continuous data involves construction of a histogram. The value of the bin with the greatest number of data points is the mode, and this value can be fine-tuned by simple interpolation with adjacent bins []. The major drawback of this method is that different modes can be obtained using different bin sizes, although some stability can be gained by using the mean of modes obtained from different bin sizes.

b. ***Direct methods***

Several new mode estimation methods that do not require density estimation have been proposed in recent years. Two related ones are the Half-Sample Mode (HSM) and Half-Range Mode (HRM), which are based on iterative bisection respectively seeking for the shortest half sample or the densest half sample. Of these two methods, HRM is commonly preferred since it has been shown to have lower bias with increasing contamination and asymmetry [80].

4.5.1.2 Bracho and Sanderson's method (1st category)

By assuming that the noise is Gaussian distributed, Bracho and Sanderson noticed that, for a noisy image with no structure, the magnitude of the intensity gradient will be Rayleight distributed. Since the Rayleight probability density function has a maximum for a value equal to σ_n this may be computed from the histogram of the gradient magnitude.

To increase robustness, the histogram is smoothed before the peak is found. If the image contains large regions with roughly uniform intensity, the image structure mainly will affect the tail of the distribution, and will not significantly affect the localization of the peak. However, the strategy may fail when most of the image is dominated by texture.

---

[103] Notice that using overlapping windows instead of blocks smoothes the pdf (the sample variance at a pixel can be seen as average between variance of neighbouring pixels). In other words, variances are correlated.

4.5.1.3 Immerkær's method (2nd category)

Immerkær presented in [94] a fast and simple method for estimating the variance of additive zero mean Gaussian noise in an image. To remove image structure, it applies to the noisy image a 3x3 linear separable filter *N*, computed as the weighted difference of two Laplacian filters, L_1 and L_2, which estimates the second derivate of the image signal. The effect of *N* is to reduce constant, planar and quadratic 3x3 facets to zero plus a linear combination of the noise. Conceptually, this is equivalent to computing the residual of a quadratic surface fitting, without explicitly computing the surface, what significantly reduces the computational complexity.

$L_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ $L_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & -4 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ $N = L_2 - 2L_1 = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$

Figure 4.16 Filter masks used by Immerkaer's noise variance estimation technique.

Once the image has been filtered to remove structure, the filtered pixel values can be used to compute the estimated noise variance, $\hat{\sigma}_1^2$, according to

$$\hat{\sigma}_1^2 = \frac{1}{\| N \|_{L^2}^2} \sum_{i \in I} (v(i) * N)^2 \; ; \; \| N \|_{L^2}^2 = 36 \qquad \hat{\sigma}_2^2 = \frac{\pi}{2 \| N \|_{L^2}^2} \left(\sum_{i \in I} | v(i) * N | \right)^2$$

where *v(i)*N* denotes the result of applying filter *N* a pixel *v(i)*.

By using the fact that, for a zero mean Gaussian random variable X, $E[X^2] = \pi/2 \, E[|X|]$, an alternative method for computing the estimated noise is $\hat{\sigma}_2^2$.

This formulation has two advantages: *i)* the summation requires no multiplications, and *ii)* the absolute deviation is more robust to the presence of outliers [93].

4.5.1.3.1 Extension

First, while the method performs quite well for a large range of noise variance values, it still overestimates in textured images or when the noise level is very low. For such situations, we propose to improve Immerkaer's solution by using alternative robust methods, such as the MAD (median absolute deviation from the median), which have not been investigated by the author (probably due to the increased computational complexity). Second, we observe that *N* is the *high-pass* equivalent of a 3x3 binomial kernel [104]. This allows extending it to larger supports and understanding its frequency behaviour: *N* results to be the narrowest (i.e. with smallest bandwidth) *high-pass* filter achievable with a 3x3 support, thus rejecting the highest amount of *low-frequency* image data while at the same time keeping *high-frequency* noise.

---

[104] A *high-pass* version is obtained from a *low-pass* smoothing kernel by multiplying each coefficient by $(-1)^{i+j}$, where *i,j* are the matrix indexes.

4.6 Proposed Approach

We propose to use NL-means as state-of-the-art edge-preserving smoothing filter to decompose an image in its *intrinsic* and *extrinsic* (be it noise or illuminance) components. This decision is twofold, theoretical and practical. First, with an extremely simple and intuitive formulation (which can in fact be interpreted as linear filtering in a higher-dimensional space), yet being strongly related to Bayesian regularization, diffusion through PDEs, Robust Statistics (M-estimation), Mean-Shift, etc., as we have shown, neighbourhood filters provide an extraordinary unifying theoretical framework. Second, from a practical point of view, NL-means has demonstrated strong superiority compared to bilateral filter in terms of visual quality and is easily extendible to colour images and cross/joint filtering.

In order to account for noise spatial and cross-channel correlation as well as signal-dependence present in real images, we further propose to respectively down-weigh RGB values by signal and noise covariance matrices[105], setting the range sigma to be a function of the estimated noise level for each image intensity level.

We observe that high-pass filters used in noise estimation methods, including Immerkaer's separable kernel, can be interpreted as a good approximation of the smallest eigenvector of the neighbourhood covariance matrix, which roughly corresponds to a high-pass filter for natural images. Then the smallest eigenvalue provides a good estimate of the noise variance. Besides serving as unifying framework, principal component analysis (PCA) may yield improved accuracy for noise estimation in textured images.

Further improvement may be achieved by looking at the histogram of the resulting high-pass sub-band: it is possible to accurately estimate noise level by fitting a Gaussian distribution to small coefficient values, as shown in the figure.

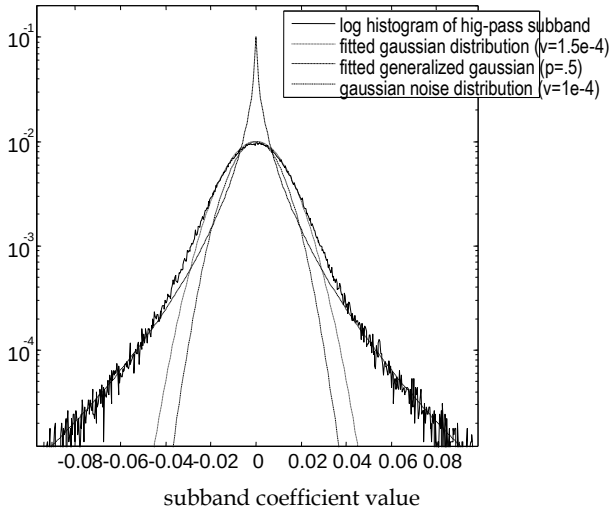


Figure 4.17 Log histogram of high-pass sub-band and comparison with noise's Gaussian distribution.

[105] Notice that noise variance estimation only provides knowledge about noise power but not about intra and between channels correlations, which have proven to be of high influence in both, quality assessment and algorithm performance. It is thus convenient to transform first the image into a new basis for which the noise is spherical (uncorrelated components with the same variance) and the signal vector density is elliptical and aligned with the axes (uncorrelated components, but with different variance each one).

4.7 Validation of results

A classical comparison receipt based on noise simulation consists of taking a good quality image, adding Gaussian white noise with known σ, and then computing the best image recovered from the noisy one by each method. A table of L^2 distances (i.e., MSE) from the restored to the original can be established. The L^2 distance does not provide a good quality assessment, since it does neither ensure that the original image features are preserved nor artifacts are not introduced, a requirement which is not usually demanded by denoising algorithms. In order to better evaluate and compare the performance of denoising algorithms, we propose using the following two principles [83]:

a. ***Preservation of original information***: since features in the residual or "*method noise*" $n(D_h,v)=v-D_hv$ are removed from v, for every denoising algorithm, the residual must be zero if the image contains no noise and should be in general an image of independent zero-mean random variables.[106] Otherwise, the residual can be filtered again and its deterministic part turned back to the image. Recent denoising methods adopted this recursive strategy to recover image information lost in *method noise* [83]. The outcome of such experiments has shown a clear cut on a wide class of denoising filters of all origins including all mentioned neighborhood filters, being the NL-means method noise the one which looks the more like a white noise.

b. ***Introduction of no artifacts:*** Because it is impossible to totally remove noise, an important question is how remnants of noise look like, since the transformation of a white noise into any correlated signal creates structure and artifacts. Only white noise is perceptually devoid of structure, as was pointed out by Attneave [76]. A denoising method must transform a white noise image into a white noise of lower variance. This requirement seems to be the best way to characterize artifact-free methods, since it eliminates any subjectivity and can be checked by mathematical arguments (e.g., Fourier analysis). These, together with experimental ones, have shown that neighborhood filters are the only ones satisfying this principle. In order to compare the latter, Buades et al. [83] have introduced a third comparison principle, namely "***statistical optimality***", which questions whether a given neighborhood filter is able or not to retrieve faithfully the neighborhood $J(i)$ of any pixel i, i.e., all and only the pixels j having the same model as i. The NL-means has been shown to best match this requirement.

Mathematical and experimental arguments haves hown that bilateral filters and NL-means are the only ones satisfying the noise to noise principle:

[106] It is much easier to evaluate whether a method noise contains some structure removed from the image or not. However, when the standard deviation is higher than the feature contrast, a visual exploration is not reliable, since image features can be masked in the residual.

i) ***Gaussian Convolution:*** The convolution with a Gauss kernel G_h is equivalent to the product in the Fourier domain with a Gauss kernel of inverse standard deviation $G_{1/h}$. Therefore, convolving the noise with a kernel reinforces the low frequencies and cancels the high ones. Thus, the filtered noise will no more be a white noise and actually shows big grains due to its prominent low frequencies.

ii) ***Wavelet Thresholding:*** Noise filtered by a wavelet thresholding is no more a white noise. The few coefficients with a magnitude larger than the threshold are spread all over the image. The pixels which do not belong to the support of one of these coefficients are set to zero. The visual result is a constant image with superposed wavelets. It is easy to prove that the denoised noise is spatially highly correlated.

iii) ***Bilateral Filter:*** For simplicity consider the case where the grey level neighborhood is an interval. Given a noise realization, the filtered value by the bilateral filter at a pixel i only depends on its value n(i) and the parameters h and ρ. The bilateral filter averages noise values at a distance from n(i) less or equal than h. Thus as the size ρ of the neighborhood increases by the law of large numbers the filtered value tends to the expectation of the Gauss distribution restricted to the interval (n(i) − h, n(i) + h). The filtered value is therefore a deterministic function of n(i) and h. Independent random variables are mapped by a deterministic function on independent variables. Thus the noise to noise requirement is asymptotically satisfied by the bilateral filter.

iv) ***NL-Means Algorithm:*** NL-means satisfies the noise to noise principle in the same extent as a classical bilateral filter. However, a mathematical statement and proof of this property are intricate and we shall skip them.

a) noisy image *b)* denoised image *c)* residual or *method noise*

Figure 4.18. Method noise of NL-means filter on a fine detailed noisy image. The original noisy image in a) has been filtered using NL-means algorithm to yield the edge-preserved smoothed image in b). Obser that, while the residual (difference between noisy and denoised images) has been scaled by 2 in order to make details more visible, resulting in image c), it looks almost like white Gaussian noise, lacking any detail or structure other than some vertical artefacts, which we suspect are the result of compression algorithms

We have proposed three principles for the comparison of denoising methods evaluating the *loss of image structure*, the *creation of artifacts*, and the *complete usage of image self-similarity*. Buades et al. have shown that: ***i)*** only wavelet thresholding methods and NL-means give an acceptable method noise; ***ii)*** neighborhood filters are the only ones to satisfy the "noise to noise" principle, ***iii)*** among them NL-means is closest to statistical optimality.

There is, however, no numerical criterion to judge the performance of a denoising algorithm in a real problem. Note that every criterion measures a different aspect of the denoising method. It is easy to show that only one criterion is not enough to make a judgement, and so one expects a good solution to have a high performance under all the criteria. The human eye is the only one able to decide if the quality of the image has been improved by the denoising method. For this reason, the non-presence of artifacts is one of the main requirements to denoising algorithms.

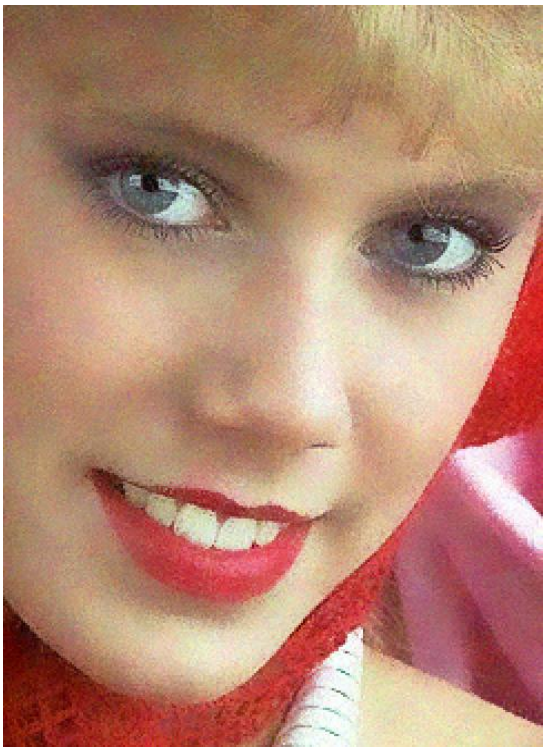
We display some denoising experiences comparing the NL-means algorithm with local smoothing filters. All experiments have been simulated by adding a Gaussian white noise of standard deviation $\sigma^2 = 0.02$ to the true image. The objective is to compare the visual quality of the restored images, the non-presence of artifacts and the correct reconstruction of edges, texture and details.

a) Original image ***b)*** Synthetic noisy image $(\sigma^2 = 0.02)$

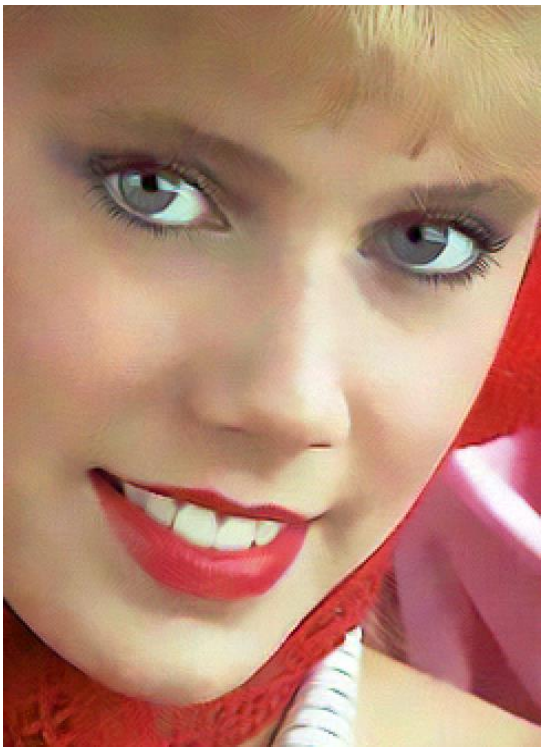
Figure 4.19. Original image and synthetic noisy image, generated by adding a Gaussian white noise of standard deviation $\sigma 2 = 0.02$ to the true image. The use of faces a test images in image denoising is commonplace because, besides being a very familiar subject, the presence of both smooth areas, such as the cheeck, as well as fine detail areas, such as hair and eyelashes, allows a proper visual evaluation of algorithms' performance in terms of detail preservation and lack of artefacts

a) Nadaraya-Watson, Gaussian *(h_s=7 pixels)*

b) Wiener filter *(h_s = 7 pixels)*



c) Bilateral filter *(h_s=7, h_r=0.56)*



d) NL-means *(h_s=7, pattern 3x3, h_r=0.045)*

Figure 4.20. Comparison of denoising algorithms' performance. We visually compare the the performance of the different algorithms presented in this chapter, from the linear (hence, *non*-edge-preserving) Nadaraya-Watson estimator using Gaussian kernel to the Non-Local means filter. Observe that in all cases, the spatial neighborhood is determined by a Gaussian kernel with a radius of 7 pixels, being the tonal neighborhood strategy what differentitates each algorithm. Both spatial and tonal scaling parameters were determined empirically for the purpose of this exercise.

4.8 Summary

This chapter has reintroduced and expanded the *edge-preserving* image smoothing framework, as a valuable digital darkroom tool for the task of simplification of visual information, with a variety of applications in computer graphics and image processing, e.g., in *image restoration*, *multi-scale tone management*, *style-transfer* or *image editing*, where it is often paramount to have the edges, or *features* in general, preserved by the image coarsening process. While the focus has been to stablish the relation with existing popular and very recent denoising techniques, the tools developed and the results obtained can easily be extended to any other application area where it is required to split an image into a (piecewise) smooth base layer, containing large scale variations in intensity, and one (or more, in the case of multi-scale decompositions) residual detail layer(s) capturing the smaller scale details in the image. Depending on the settings and the application, this small-scale component can be interpreted as *noise* or *texture*. In general, each of the resultant layers may be regarded either as an *intrinsic* (e.g., reflectance) or *extrinsic* (e.g., illuminance) component of the image, thus manipulated separately in various ways (e.g., emphasized, compressed, or even discarded) and possibly recombined to yield the final result. As such, edge-preserving smoothing filters are a key tool in our mission to provide a system for automated image improvement.

The regularity in data fundamentally distinguishes itself from random noise and describing it in generic, yet powerful, ways has traditionally been one of the key problems in signal processing. Instead of incorporating a priori explicit information into the image model itself, what lacks the generality to be easily applied to diverse image collections, we looked for statistical-inference methodologies that allow us to implicitly learn the underlying structure via data-driven strategies, thus providing powerful tools in formulating unsupervised adaptive image-processing methods. In fact, the very recent methods here presented suggest that it is possible to take advantage of an image model learned from the observed image itself. More specifically, these denoising methods attempt to learn the statistical relationship between the image values in a window around a pixel and the pixel value at the window centre.

The material presented in this chapter builds from previous research, highlighting the importance of integrating very diverse classical and recent approaches, from *regularisation* techniques, *nonlinear diffusio*n filtering and *adaptive smoothing* to *mode filtering*, *mean shift*, *kernel regression*, *M-estimators* from *robust statistics*, *Bilateral Filter* and *non-local means* filtering. While these methods originate in very different frameworks, excelling from certain interesting angle but also inevitably subject to their limitations and applicability, they all can be cast into the unified framework of *energy (functional) minimisation*, where the output of the algorithm or the *'estimate'* is a global or local minimum of a *'loss function'*. This combines (possibly nonlocal) *data fidelity* and (possibly nonlocal) *prior smoothness* terms, where the mutual influence of image pixels is respectively

controlled by (robust) weighting functions depending on the tonal and spatial distances.

We observe also the equivalence to applying classical linear procedures in a higher-dimensional space, using a single constant weight function based on the Euclidean distance defined on the joint spatial-tonal domain $S \times R$. In fact, assuming the pixel intensity function is smooth as a function of the noisy patches (features), then edge-preserving smoothing can be achieved by running the heat equation (i.e., linear convolution) in the feature-space and projecting back the obtained solution to the original image domain, if needed. Several state-of-the-art denoising methods result from this formulation through different choices of the features: from pixel intensity and local neighbourhood or *'patch'* to filter responses such as wavelets. It is also noticed that, the less the noise changes the geometry of the selected feature-space, i.e. the distance and ordering between patches, the more robust the smoothing will be, provided the noise is not correlated neither with the image nor with itself.

Covering a larger number of methods and including them all into a single, unified framework has several advantages. First, it explicitly shows all the freedom in selecting the penaliser type, the parameters, and the balance between smoothness and data terms. Second, it makes explicit what assumptions are needed to derive known methods, clarifies their relations and brings new insights, contributing to a better understanding and opening the way to novel techniques to combine the advantages of known filters and fit the particular properties of the data and noise.

With the intention to provide self-content tools, we include a more realistic signal-dependent noise model and propose an improved noise level estimation method based on a Principal Component Analysis generalization of state-of-the-art methods. It will allow us to assess the impact on image quality due to the presence of noise, select the appropriate Image Processing strategy, set initial values of tuning parameters and eventually adapt the smoothing degree to the local noise characteristics.

Future work

We observe that keeping the noise estimator as a separate module, which may be replaced with better technique if one becomes available, may however yield to suboptimal solutions. Ideally, the processes of noise estimation and denoising should be intimately merged in one. Moreover, the Bayesian framework provides a formal way for choosing appropriate tonal kernels for the data and smoothness terms, restricting the parameter space depending on the noise. Studying other types of noise and the properties of the signal to recover will lead to different criteria for selecting the penalisers. In Section 4.9.1 *Sources of Noise* we provide a brief summarize the most important characteristics of relevant noises present in digital images.

4.9 Appendix

4.9.1 Sources of Noise

Noise occurs in images for many reasons. It may be introduced by the medium through which the image is created (e.g., random absorption or scatter effects), by the recording medium (sensor noise), by measurement errors due to the limited accuracy of the recording system, by quantization of the data for digital storage and, in general, by any transmission or communication process through a noisy channel (storage may also be modeled as such).

In what follows we summarize the most important characteristics of relevant noises present in digital images.

4.9.1.1 Photographic Noise

Many digital images still have in their origin a photochemical process, where typically millions of tiny silver halide grains change their chemical properties when exposed to light. Common assumptions when modeling this process are (1) grains are uniform in size and character and (2) the probability of each one changing its appearance[107], $\boldsymbol{p}$, is proportional to the number of striking/incident photons. Then the number of grains that change, $\boldsymbol{N}$, among the $\boldsymbol{L}$ grains contained in a given area $\boldsymbol{A}$, is binomial

$$\Pr(N = k) = \binom{L}{k} p^k (1-p)^{L-k} \qquad (4.47)$$

For $\boldsymbol{L}$ large enough (which is the typical case), this probability is well approximated by a Poisson distribution when $\boldsymbol{p}$ is small but $\boldsymbol{\lambda=Np=E[N]}$ moderate, and by a Gaussian distribution (with mean $\boldsymbol{Lp}$ and variance $\boldsymbol{Lp(1-p)}$) when p is larger. This variance is maximized when $\boldsymbol{p=0.5}$ [78]

This stochastic nature of the photochemical image formation process results in the so-called ***photographic grain*** noise, $\boldsymbol{N_g}$, a well-known characteristic of photographic films which limits the effective magnification one can obtain from a photograph.

[107] This is done by becoming metallic silver. In the developing process, the unchanged grains are washed out.

4.9.1.2 CCD Imaging

CCD devices show three main kinds of noise [121]. From image irradiance at sensor plane to digitalized gray-level values, these are: shot or photon, thermal or dark current and read-out noises. In what follows, these are briefly described.

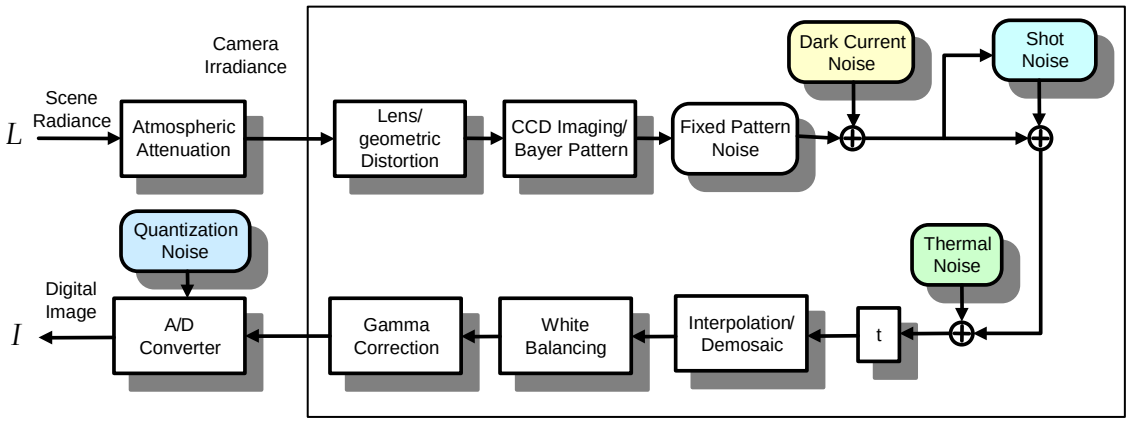


Figure 21. CCD camera imaging pipeline, reproduced from [103].

A CCD is used to measure the spatial distribution of light incident on a thin wafer of silicon. The measurement process relies on the fact that when a photon strikes silicon, an electron-hole pair is generated (this is known as the photoelectric principle, discovered first by Hertz and later explained by Einstein, for which he was awarded the Nobel Prize). These electrons are "captured" in a well and, after some time, counted by a "read out" device. The number of electrons counted, N, can be written as

$$N=N_I+N_{th}+N_{ro}$$

where N_I is the number of electrons due to the image (i.e. ***photoelectrons***), N_{th} the number due to thermal noise, and N_{ro} the number due to read out effects.

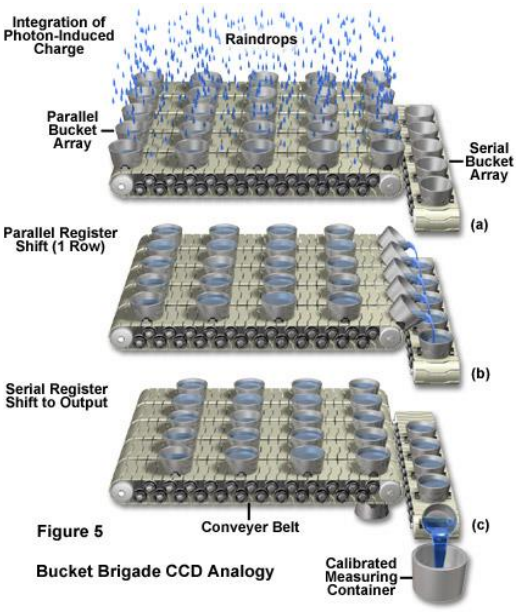


Figure 42. Bucket Brigade CCD Analogy: The operation of a CCD is often compared to measuring the spatial distribution of rainfall over a field by placing an array of buckets on it [114].

we will assume that: *i)* the number of electrons collected at each site is independent of the number of electrons collected at other sites; and *ii)* our imaging system is configured to avoid overilluminating individual collection sites. Thus, saturation and blooming effects are not considered. Blooming occurs when a CCD well is filled and additional photoelectrons spill over into adjacent CCD wells.

4.9.1.3 Shot or Photon Noise

This characterizes the uncertainty in the number of photoelectrons stored at a collection site, due to the random quantum (discrete) nature of light. Since cannot be eliminated, introduces a fundamental limitation.

In essence, most image acquisition devices are photon counters (modern CCDs are sensitive enough to count individual photons) [78]. The probability distribution for k photons in an observation window length of Δt seconds is known to be Poisson[108], with expected value (mean) proportional to the incident image intensity

$$\Pr(N = k) = \frac{e^{-\rho\Delta t}(\rho\Delta t)^k}{k!} \quad (4.48)$$

$$(\mu = \rho\Delta t = \sigma^2) \qquad SNR = 10\log_{10}(\rho\Delta t)\ dB$$

where ρ is the rate or intensity parameter measured in photons per second. Poisson processes have the following important (for imaging) properties:

a) The variance is equal to the average number of events.

b) Non-overlapping exposures are statistically independent events [106].

c) The Poisson process is additive: the sum of two independent Poisson-distributed RVs with means μ_1 and μ_2 is also Poisson distributed with mean and variance $\mu_1+\mu_2$.

Property *(a)* yields a signal-dependent noise model. Photon noise is larger in bright parts than in dark ones, an effect which is reduced and sometimes even reversed by gamma-correction. Properties *(b)* and *(c)* together mean that we can reduce the noise variance by averaging many images captured with the same sensor at different times. Observe from eq. that the only way to increase the SNR is by means of capturing more photons, what can be achieved by increasing scene luminance, lens aperture and/or exposure time.

However, CCD arrays saturate due to a finite well capacity, C, proportional to pixel size and well depth[109]. This means that the maximum SNR per pixel attainable by a CCD camera is given by [121]

$$SNR = 10\log_{10}(C)\ dB$$

Typical pixel sizes are between 6.8 x 6.8 and 23.0 x 23.0 μm², yielding capacities between 32,000 and 320,000 electrons per site, resulting in raw (i.e. before processing) SNR of 45-55dB just by the underlying Poisson process. This

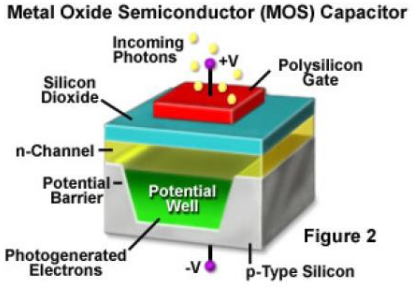


Figure 23 Metal Oxide Semiconductor (MOS) Capacitor.

[108] A random process in which we count on average $\lambda\Delta t$ events is known as a ***Poisson process.***

[109] While different sensors may have different pixel size, well depth, as a consequence of CCD technology, is constant at about 700 $e^-/\mu m^2$ [121].

may worsen because of noise amplification by further processing (e.g., gamma-correction) as well as other sources of noise.

4.9.1.4 Thermal or Dark Current Noise

Due to thermal vibrations, (energy) electrons can be freed from CCD material even in absence of light (hence the term ***dark current***). In fact, these cannot be distinguished from true photoelectrons. Thermal noise is proportional to exposure time and Poisson-distributed, with the rate parameter being an increasing function of the temperature. Thus, it can be reduced by cooling the sensor and using higher shutter speeds (at the price of more scene illuminance and/or lens aperture).

Nevertheless, it is typically assumed to be AWGN. The zero-mean property is due to the subtraction from the raw image of a *'dark frame'*, obtained by averaging several images taken with the shutter closed, but with the same shutter speed and sensor temperature [83].

4.9.1.5 Read out Noise

Electronic and signal independent, can be assumed to have zero mean by the subtraction from the raw image of a 'bias frame', obtained by averaging several images taken with the shutter closed and a zero-exposure time (so that any signal measured is due to read out effects) [78].

Notice that noise is especially important in the blue channel, where more amplification is required due to the reduced sensitivity of sensors to short wavelengths.

4.9.1.6 Quantization Noise

In order to be numerically processed, images must be first converted to a digital representation, what necessarily conveys a quantization process. In the CCD camera imaging pipeline this takes place in the analog-to-digital converter (ADC), commonly as a final step before storage (assuming that no digital image processing occurs before).

Rounding errors result in a ***quantization noise***, N_q, which for a small number of (quantization) levels (L<16), is signal-dependent (i.e., in an image of the noise, signal features can be discerned), spatially correlated and not uniformly distributed. However, for larger L, which is the common case (images are typically digitalized with b=8 or 16 bits, yielding L=256 and 65,536 levels, respectively), N_q can be reasonably assumed to by uniform distributed within the interval $[-\frac{1}{2}q, \frac{1}{2}q]$, where q is the quantization step:

$$N_q \sim U(-\tfrac{1}{2}q, \tfrac{1}{2}q) \qquad (\mu = 0;\ \sigma^2 = q^2/12)$$

Thus, each additional bit in the quantizer results in an SNR increase of 6dB SNR=6b+11 (dB). For b=8 bits, SNR=59 dB.

4.9.2 Notation

- $\mathcal{E}$ Error
- $\mathcal{C}$ Reliability or certainty
- ψ (Empirical) Influence function
- ρ (Robust) Error norm. $\boxed{\psi(x) = \rho'(x) = g(x) \cdot x}$
- σ Scale parameter.
- g_σ, w_σ Weight with σ bandwidth
- G_h Gaussian kernel of standard deviation *h*.
- K_h Kernel with scale parameter *h*.
- i, j Index
- $\mathbf{x} = (x_i)_{i=1,\ldots,N}$ Vector (by default considered as column vector, like in MATLAB)
- $\|\mathbf{x}\|$ L_2 norm.
- $\|\mathbf{x}\|_p$ L_p norm.
- $\mathbf{A}$ Matrix
- $\mathbf{I}$ Identity matrix
- $\mathbf{x}^T, \mathbf{A}^T$ Vector, Matrix transpose
- Ω_S image or spatial domain
- Ω_R radiometric, feature or tonal domain
- $\mathrm{f} : \Omega_S \to \Omega_R$ Image defined as a function
- V, U, N Observed, Original (or "true"), and Noise random variables
- $\mathbf{v}, \mathbf{u}, \mathbf{n} \in \Re$ Observed, Original (or "true"), and Noise singular vectors (realizations)
- $v_i, u_i, n_i \in \Omega_R$ Observed, Original (or "true"), and Noise value of sample *i*.
- $\vec{x}_i = [x_1, x_2]_i^T \in \Omega_S$ Position of sample *i*, or *'pixel'*, when samples are in a 2D grid.
- $\vec{U}_i = \left\{ \frac{\vec{x}_i}{\sigma_s}, \frac{\vec{u}_i}{\sigma_R} \right\}$ Generalized intensity
- $\mathbf{C}_{\mathbf{xy}}, \mathbf{S}_{\mathbf{xy}}$ Covariance matrix, Cov(**x**,**y**) or Cov(X,Y)
- $s_{\mathbf{x}}$ Variance, Var(**x**)
- $\nabla \mathbf{u}$ Gradient of **u**
- Δ Laplacian $\Delta u = \nabla \cdot \nabla u = \nabla^2 u$
- f' Derivate

- $\beta(i)$ Neighbourhood of sample *i*
- i~j $\Leftrightarrow i \in \beta_4(j) \Leftrightarrow j \in \beta_4(i)$
- $E(\mathbf{u}), \overline{\mathbf{u}}$ Mean of **u**.
- $\mathrm{J}(\mathbf{u})$ Functional (energy)
- $\propto$ Proportional to
- $p_x(x;\vartheta)$ pdf parametrized by ϑ
- $\hat{\vartheta}$ estimated ϑ
- $p_{x|y}(x \mid y)$ Conditional pdf
- $\mathcal{N}(\mu,\sigma)$ Normal or Gaussian pdf with mean μ and std. dev. σ
- λ eigenvalue
- $\log$ natural logarithm
- MAD Median Absolute Deviation
- MAP Maximum A Posteriori
- ML Maximum Likelihood
- MSE Mean Squared Error
- MMSE Minimum MSE
- LS Least Squares

REFERENCES

[75] ARGENTI, Fabrizio; TORRICELLI, Gionatan; ALPARONE, Luciano. MMSE filtering of generalized signal-dependent noise in spatial and shift-invariant wavelet domains. *Signal Processing*, 2006, vol. 86, no 8, p. 2056-2066.

[76] ATTNEAVE, Fred. Some informational aspects of visual perception. *Psychological review*, 1954, vol. 61, no 3, p. 183.

[77] AUBERT, Gilles; KORNPROBST, Pierre. *Mathematical problems in image processing: partial differential equations and the calculus of variations*. Springer Science & Business Media, 2006.

[78] BOVIK, A. C. Handbook of Image and Video Processing. 2nd ed. Elsevier Academic Press, 2005

[79] BARASH, Danny; COMANICIU, Dorin. A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift. *Image and Vision Computing*, 2004, vol. 22, no 1, p. 73-81.

[80] BICKEL, David R. Robust estimators of the mode and skewness of continuous data. *Computational statistics & data analysis*, 2002, vol. 39, no 2, p. 153-163.

[81] BLACK, Michael J., et al. Robust anisotropic diffusion. *IEEE Transactions on image processing*, 1998, vol. 7, no 3, p. 421-432.

[82] BROX, Thomas; KLEINSCHMIDT, Oliver; CREMERS, Daniel. Efficient nonlocal means for denoising of textural patterns. *IEEE Transactions on Image Processing*, 2008, vol. 17, no 7, p. 1083-1092.

[83] BUADES, Antoni; COLL, Bartomeu; MOREL, J.-M. A non-local algorithm for image denoising. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005. p. 60-65.

[84] CHOUDHURY, Prasun; TUMBLIN, Jack. The trilateral filter for high contrast images and meshes. *ACM SIGGRAPH 2005 Courses*. ACM, 2005. p. 5.

[85] DOOLEY, R.P.; SHAW, R.: Noise Perception in Electrophotography, J. APPL. PHOTOGR. ENG., vol. 5, no. 4

[86] DURAND, Frédo; DORSEY, Julie. Fast bilateral filtering for the display of high-dynamic-range images. *ACM transactions on graphics (TOG)*. ACM, 2002. p. 257-266.

[87] ELAD, Michael, and MICHAL Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing* 15, no. 12 (2006): 3736-3745.

[88] CHU, C. K., et al. Edge-preserving smoothers for image processing. *Journal of the American Statistical Association*, 1998, vol. 93, no 442, p. 526-541.

[89] COMANICIU, Dorin; MEER, Peter. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 2002, vol. 24, no 5, p. 603-619.

[90] FUKUNAGA, Keinosuke; HOSTETLER, Larry. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 1975, vol. 21, no 1, p. 32-40.

[91] GEMAN S. and GEMAN D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions On Pattern Analysis and Machine Intelligence*. Vol. 6, No. 6, June 1984.

[92] HAMPEL, Frank R., et al. *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.

[93] HUBER, Peter J. Robust statistics. *International Encyclopedia of Statistical Science*. Springer Berlin Heidelberg, 2011. p. 1248-1251.

[94] IMMERKAER, John. Fast noise variance estimation. *Computer vision and image understanding*, 1996, vol. 64, no 2, p. 300-302.

[95] KAYARGADDE, Vishwakumara; MARTENS, Jean-Bernard. An objective measure for perceived noise. *Signal Processing*, 1996, vol. 49, no 3, p. 187-206.

[96] KERVRANN, Charles; BOULANGER, Jérôme; COUPÉ, Pierrick. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. *Scale Space and Variational Methods in Computer Vision*, 2007, p. 520-532.

[97] KINDERMANN, Stefan; OSHER, Stanley; JONES, Peter W. Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling & Simulation*, 2005, vol. 4, no 4, p. 1091-1115.

[98] KOENDERINK, Jan J.; VAN DOORN, Andrea J. The structure of locally orderless images. *International Journal of Computer Vision*, 1999, vol. 31, no 2, p. 159-168.

[99] KNUTSSON, Hans; WESTIN, C.-F. Normalized and differential convolution. *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*. IEEE, 1993. p. 515-523.

[100] LEE, Yong; KASSAM, S. Generalized median filtering and related nonlinear filtering techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1985, vol. 33, no 3, p. 672-683.

[101] LEE, J. S. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2, pp. 165–168, 1980.

[102] LEE, J. S.; HOPPEL, K. Noise modeling and estimation of remotely-sensed images. *Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th Canadian Symposium on Remote Sensing., 1989 International*. IEEE, 1989. p. 1005-1008.

[103] LIU, Ce, et al. Noise estimation from a single image. *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. IEEE, 2006. p. 901-908.

[104] MRÁZEK, Pavel; WEICKERT, Joachim; BRUHN, Andres. On robust estimation and smoothing with spatial and tonal kernels. *Geometric properties for incomplete data*. Springer Netherlands, 2006. p. 335-352.

[105] OLSEN, Soeren I. Estimation of noise in images: An evaluation. *CVGIP: Graphical Models and Image Processing*, 1993, vol. 55, no 4, p. 319-323.

[106] PAPOULIS, A. Probability, Random Variables, and Stochastic Processes. MacGraw-Hill, New York, incl. edn. 1991.

[107] PARIS, Sylvain, et al. A gentle introduction to bilateral filtering and its applications. *ACM SIGGRAPH 2007 courses*. ACM, 2007. p. 1.

[108] PERONA, P. and MALIK, J. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions On Pattern Analysis and Machine Intelligence*. Vol. 12, No. 7, July 1990.

[109] PHAM, Tuan Q.; VAN VLIET, Lucas J.; SCHUTTE, Klamer. Robust fusion of irregularly sampled data using adaptive normalized convolution. *EURASIP Journal on Advances in Signal Processing*, 2006, vol. 2006, no 1, p. 083268

[110] PORTILLA, Javier, et al. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image processing*, 2003, vol. 12, no 11, p. 1338-1351.

[111] RANK, K.; LENDL, M.; UNBEHAUEN, R. Estimation of image noise variance. *IEE Proceedings-Vision, Image and Signal Processing*, 1999, vol. 146, no 2, p. 80-84.

[112] RAPHAN, Martin; SIMONCELLI, Eero P. Learning to be Bayesian without supervision. *Advances in neural information processing systems*. 2007. p. 1145-1152.

[113] SOCHEN, Nir; KIMMEL, Ron; MALLADI, Ravi. A general framework for low level vision. *IEEE transactions on image processing*, 1998, vol. 7, no 3, p. 310-318.

[114] SPRING, K. R.; FELLERS, T. J.; DAVIDSON, M. W. Introduction to charge-coupled devices. *MicroscopyU: the source for microscopy education Retrieved June*, 2010, vol. 2, p. 2010.

[115] TAKEDA H., FARSIU S. and MALINFAR P. Kernel Regression for Image processing and Reconstruction. *IEEE Transactions on Image Processing*, Vol. 16, No. 2, February 2007.

[116] TOMASI, Carlo; MANDUCHI, Roberto. Bilateral filtering for gray and colour images. *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998. p. 839-846.

[117] VAN DEN BOOMGAARD, Rein; VAN DE WEIJER, Joost. On the

equivalence of local-mode finding, robust estimation and mean-shift analysis as used in early vision tasks. *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. IEEE, 2002. p. 927-930.

[118] WINKLER, Gerhard, et al. Noise reduction in images: some recent edge-preserving methods. 1998.

[119] WINKLER, Gerhard. *Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction*. Springer Science & Business Media, 2012.

[120] WEISSTEIN, Eric W. "Robust Estimation." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/RobustEstimation.html

[121] YOUNG, Ian T.; GERBRANDS, Jan J.; VAN VLIET, Lucas J. *Fundamentals of image processing*. Delft: Delft University of Technology, 1998.

[122] Recent Trends in Denoising Tutorial. Available at http://www.stanford.edu/~slansel/tutorial/

[123] International Standard Organisation, *Photography – Electronic still-picture imaging – Noise measurements* **ISO15739:2003(E)**, Geneva, 2003.

Chapter 5

TONE REPRODUCTION

Every light is a shade, compared to the higher lights,
till you come to the sun; and every shade is a light,
compared to the deeper shades, till you come to the night.
-John Ruskin (1879)

Poor management of light (under- or over-exposed areas, light behind the main character, etc.) is the single most-commonly-cited reason for rejecting photographs. This is why camera factures have developed sophisticated exposure-metering systems. Still, a very common problem in photographic Image Enhancement (in its simplest version) is dealing with bright/dark areas which should be brighten/darken. The problem is originally caused by the fact that image capture devices do not have the adaptation capacity available in the HVS, which makes possible to face scenes with a huge amount of contrast.

Inspired in visual arts and perception literature, this chapter turns to the other extreme of the image pipeline and reviews *tone reproduction* algorithms that mimic some characteristics of the human visual system, in particular, *colour constancy* and *lightness constancy*, to discount the illuminant or, equivalently, relight the image, resulting in a more preferred image, i.e., of higher quality. This puts them on common ground with edge-preserving smoother and IQ.

5.1 Tone reproduction overview

Image *tone reproduction* refers to the operation in the imaging pipeline that maps scene luminance levels, as captured by an acquisition device (e.g., a digital camera), to luminance or density levels for display on an output device (e.g., a computer monitor o printer). Tone reproduction (a.k.a as *tone mapping*) is needed to ensure that the vast range of luminance found in a real-world scene is conveyed into the luminance range that can be produced by a given display device, while at the same time producing an image that looks satisfactory with respect to the original scene, in terms of subjective preference, appearance match and information preservation.

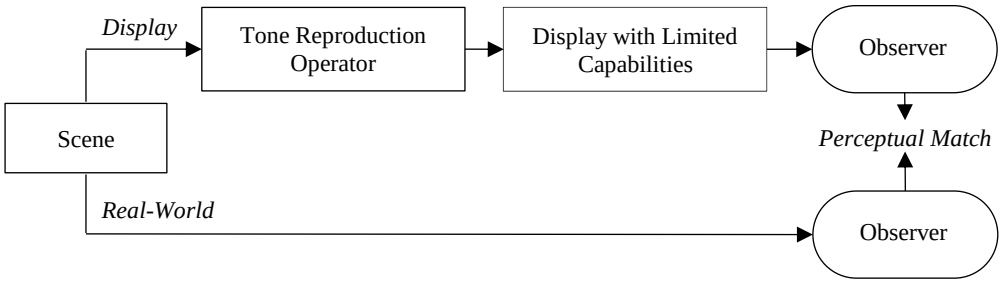


Figure 5.1. Location of tone reproduction operators in the image reproduction pipeline. The quality of the reproduction is evaluated using visual observers (usually mere mathematical models).

This is not an easy task because the displaying device or process introduces several constraints in terms of limited luminance dynamic range, narrower field of view, different observation context and adaptation level, that lead to inconsistencies in perception when viewing real world scene versus their reproductions on a display device [135][149]. The problem of tone reproduction has been around for a long time and several sensation-preserving conversions for display (better known in photography, printing, and television as *tone reproduction methods* [153]), have already been proposed to overcome the mentioned constraints.

5.1.1 Goals

The development of each of these methods was, however, driven to a great extent by requirements of a particular application and so far, a universal tone mapping algorithm is not available. For instance, in consumer imaging and commercial photography, producing just nice-looking (e.g., saturated and crispy) images is usually the main goal. Taking into account subjective preferences allows the image to look as pleasing as possible to the viewer. This is referred to as the *aesthetical* approach. In general, however, for accurate analysis and comparison with reality, the display image should bear as close a resemblance to the original scene as possible.

On the one hand, the classical *perceptual* approach bases such a resemblance on *naturalness*, usually an implicit goal in consumer imaging and realistic rendering applications, where the displayed image and the original scene should evoke the same subjective visual experience (i.e., an *appearance match*). In general,

this means that the overall impression of ***brightness***, ***contrast***, and ***colour*** should be reproduced. For example, a scene viewed at night would be represented blurred and nearly monochromatic due to scotopic vision. On the other hand, if it is important to understand the fine details or the structure of the visible lines in the result, i.e., the content of the image, the same scene would be represented with full detail. This emphasis on information preservation is the goal of the so-called *cognitive* approach, most often requested in medical and scientific imaging and archiving, where the emphasis is on *usefulness* of the conveyed visual information rather than naturalness. Summarizing, tone mapping can be profiled to achieve three different goals: *pleasantness*, *naturalness* and *usefulness*. In general, any given tone mapping operator will realize a mixture of these three approaches, with a different weighting given to each [129].

Despite the vast diversity of natural and synthetic images and the possible luminance values inaccuracy found in photographs , a robust method is in most cases expected to: 1) provide consistent results, avoiding undesirable artefacts such as perceivable contrast reversals and halos along high contrast edges; 2) to be automatic with few intuitive parameters that provide possibility for fine tuning; and 3) address different capabilities of display devices and potential differences in observation conditions [145].

5.1.2 Proposed approach

In the classical ***perceptual*** approach, quality is understood as realistic visual appearance, i.e. perceptual fidelity to the observed scene. For example, a scene viewed at night would be represented blurred and nearly monochromatic due to scotopic vision. However, if it is important to understand the fine details or the structure, i.e., the content of the image, the same scene would be represented with full detail, which would be called the ***cognitive*** approach. If the goal is only the pleasant appearance of the image, we speak about an ***aesthetical*** approach.

Tone mapping operators have generally met one of these criteria at the expense of the other. For example, some preserve the visibility of objects while changing the impression of contrast, while others preserve the overall impression of brightness at the expense of visibility. Not only preserving, but also enhancing detail visibility. In general, any given rendition will realize a mixture of these three approaches, with a different weighting given to each. This thesis focuses on **perceptual** approach, since it is the most flexible one: we can tune it emphasizing usefulness or pleasantness over naturalness.

Perceptual tone reproduction process tries to simulate the human vision process and model the tone mapping operator accordingly. As such, it is described in terms of physical processes in the display system and psychophysical processes in the hypothetical scene and display viewers that affect the fidelity of the displayed image to the scene [134]. In [153], Tumbling and Rushmeier introduced a general framework providing the theoretical basis for perceptual tone reproduction in the context of computer graphics, where

visual models[110] are used to relate the perceptual responses of a scene observer O_w to the responses of the display observer O_d in order to specify a mapping that produces a perceptual match between the scene and the display, in spite of the limited capabilities of the display device, as illustrated in Figure 5.2. Since their work, there has been a growing interest in the theme and many tone reproduction algorithms (also referred to as *tone reproduction operators* –TMOs–) have been proposed. For some very simple visual observer models, concatenating real-world and inverse display observers yields a new observer, i.e., $O_d^{-1}(O_w)=O_x$. This explains why several TMOs just take the form of a visual appearance model.

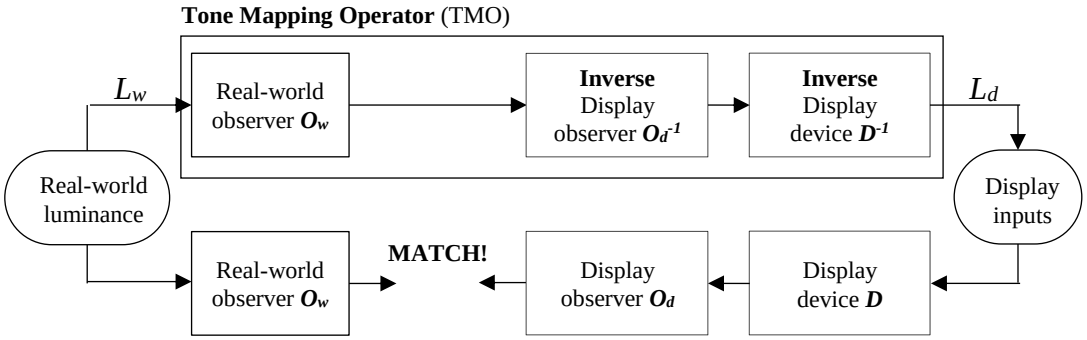


Figure 5.2. Tone reproduction as an optimization problem. A simple Tone Reproduction/Mapping Operator (TMO) may be obtained by concatenating a real-world observer model, and inverse display observer model, and an inverse display device model, i.e., $TRO = D^{-1}(O_d^{-1}(O_w))$. This procedure does not "undo" the former processes, since the visual models differ for the original scene and the display. Depending on the reproduction intent, the TMO may also include some image (e.g., detail) enhancement. Adapted from [153].

According to [143], the two criteria most common, neutral and important for perceptual tone reproduction are: 1) *naturalness* (i.e., viewing the image produces a subjective experience that corresponds with viewing the real scene) and 2) *visibility reproduction* (i.e., you can see an object in the real scene if and only if you can see it in the display). The latter relates to the concept of *usefulness*.

5.1.3 Relation to image enhancement

Researchers and practitioners in several areas encounter problems similar to that of tone mapping. In image processing, it is often desirable to increase image contrast without introducing artefacts and a full range of enhancement techniques has been developed [17][18][21]. In photography, dodging and burning techniques are popular to address the problem of obtaining a good-looking image of a high contrast scene [44]. However, while some of these techniques served as a starting point for tone mapping procedures [149], there are two fundamental differences between image enhancement and tone mapping [143]. First, maintaining perceptual fidelity is usually of no importance to an

[110] *Visual observers* are mathematical models of the HVS that account for several appearance phenomena, such as colour and light adaptation, while converting luminance values to perceived brightness images, thus providing a theoretical basis for perceptual tone reproduction. The *real-world* observer corresponds to someone immersed in the environment, and the *display* observer to someone viewing the display device.

image enhancement technique. Second, tone mapping procedure deals with undistorted world luminances while image processing algorithm is already presented with a limited dynamic range input.

Our proposal here is to get inspiration from perception models, but make no claim about their direct applicability. Specifically, we will assume that we are either provided or have the means to get a "RAW linear image", in which pixel values represent not the absolute but the relative scene luminance at each point, normalized by maximum scene luminance, a factor that will be assumed to be unknown. These are plausible assumptions, given the state-of-the-art of capture devices, which allow to virtually obtain any dynamic range.

5.1.4 Relation to photography, TV and art

Tone mapping was developed for use in television and photography, but its origins can be seen in the field of art where artists make use of a limited palette to depict high contrast scenes. It takes advantage of the fact that the HVS has a greater sensitivity to relative rather than absolute luminance levels [155]. Tone reproduction is already used extensively to good effect in photography and television [138], and some of the methods used in computer graphics have been inspired and influenced by techniques in these media.

Much of this work was conducted in the context of film-based photography, where practical considerations limited attention to global tone-mapping methods in which a single tone-mapping curve was applied to the entire image.[111] With the advent of digital imaging, a wider range of tone-mapping algorithms become of practical interest. Television and film systems have roughly sigmoid responses to light when plotted on log/log axes, as in Figure 5.3, and both use similar equations and nomenclature.

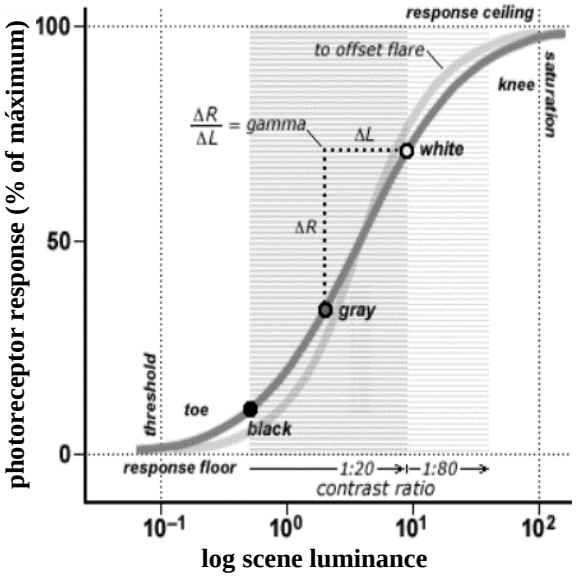


Figure 5.3. Generic *characteristic curve*, fundamentally defined by the density response of the imaging medium to light, which ranges from the zero-response baseline or response floor to a maximum response ceiling. Outside these limits, luminance variations are imageless. The threshold stimulus is the smallest light energy necessary to produce a just noticeable shift from the baseline (zero) response, and the saturation stimulus is the light energy that produces a response indistinguishable from the maximum possible. In photographic film, the response is the proportion or density of dye or silver halide crystals converted by light (from 0% to 100%); in the human eye, it is the total response range of the whole retina. Reproduced from [144], after [138].

[111] In film photography, it is not practical to automatically adjust the tone-mapping curve between images at separate locations within an image, since the shape of these curves is governed by physical characteristics of the emulsions and film-development process.

5.2 Classification and terminology

5.2.1 Classification based on spatial processing: *global* vs. *local*

Tone reproduction algorithms can be broadly classified by spatial processing techniques into two categories: *spatially uniform* (also known as *single-scale* or *global*) and *spatially varying* (also known as *multi-scale* or *local*) [130] [131][134][149]. Once the optimal transformation has been estimated according to the particular image, spatially uniform operators apply it to every pixel, regardless of the value of surrounding pixels in the image. Those techniques are simple and fast (since they can be implemented using look-up-tables), but they can cause a loss of contrast. Conversely, spatially varying operators change the parameters of the transformation according to the local features of the image. Those algorithms are more complicated than the global ones, they can show artefacts (e.g. halo effect and ringing), the output can look un-realistic, but they can provide the best performance, since the human vision is mainly sensitive to local contrast.

While not included here, temporal differences (such as adaptation over time) have also been studied under a separate category of *time dependent* tone reproduction operators.

In previous years, reviews of tone reproduction operators have been carried out with direct attention to performance comparison. Instead, this section aims to provide an overview of the conceptually most relevant approaches that have been published to date. Both, mathematical formulations and performance comparison can be found in deeper studies such as the one by Devlin in [130].

5.2.2 Terminology

Throughout the rest of this chapter, the following terminology will be used:

L	Luminance (the quantity of light in the visible range)
B	Brightness (the subjective impression of the viewer)
O	Observer. Mathematical model relating B and L: $B=O(L)$
w	real-world
d	display
a	adaptation
avg	average or mean
L'	normalized luminance, a dimensionless value in the range [0…1]. Observe that L_{min} will be in general different from 0 and $L'=1$ corresponds to L_{max}. We may, however, assume for the sake of simplicity that L has been normalized, so that $L'_{max} = 1$ and $L'_{min} = L_{min} / L_{max} \approx 0$.

5.3 Spatially Uniform techniques

5.3.1 Concept

Also called *global* operators, as they apply the same global adaptation for the whole image. Thus, the mapping function is the same for the whole image as well. The relation between input and output luminance values produced by a spatially uniform tone-mapping algorithm is called a tone reproduction *curve* (TRC).

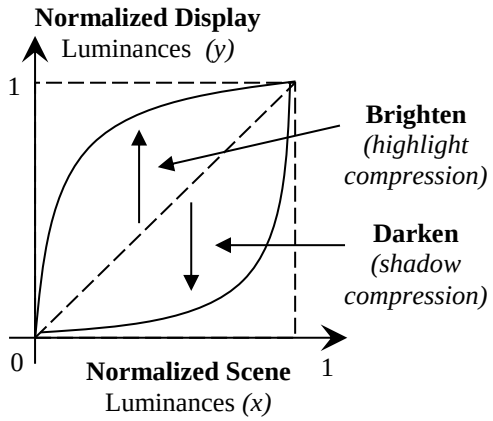


Figure 5.4. Usage of tone-mapping curve to either compress highlights or shadows.

Tone reproduction curves compress the dynamic range by defining a function that maps the original input intensities into a narrower range of display intensities. If the image input intensities are I, and the display intensities are n, then the tone reproduction curve is a function, $n = f(L)$, where $f(L)$ is a one-to-one monotonic mapping, i.e. for increasing luminance values in the scene non-decreasing luminance values of the display device will be assigned. This is required in order to avoid gradient reversals. An example of a tone reproduction curve is illustrated below.

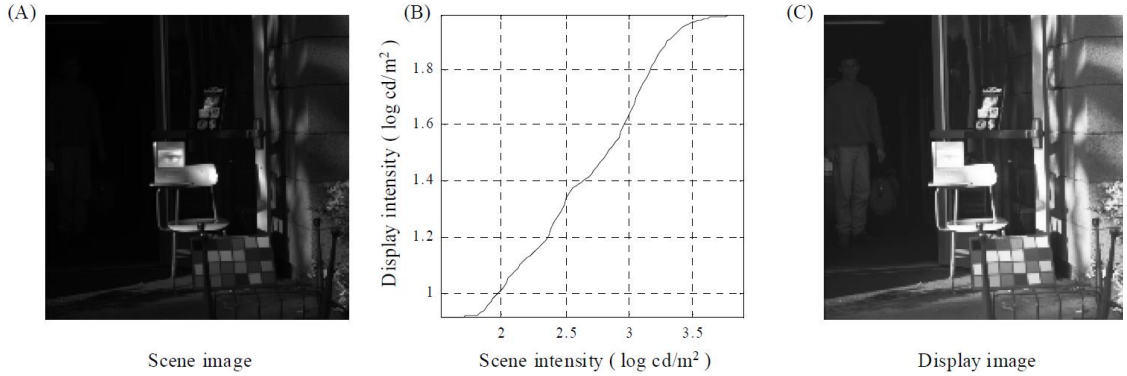


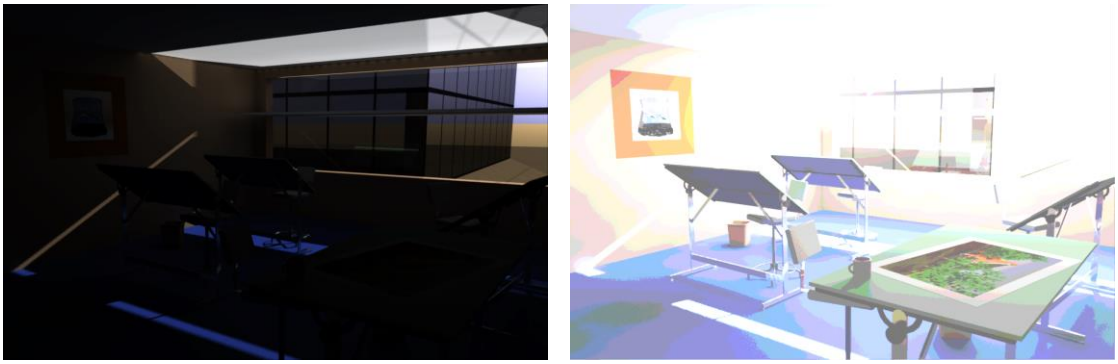
Figure 5.5. A tone reproduction curve is used to reduce dynamic range. (A) The original image. (B) The tone reproduction curve. TRCs are one-to-one and monotonic mappings. (C) The resulting image after the dynamic range has been compressed. Reproduced from [131].

More sophisticated global tone mapping methods vary the function parameters depending on global characteristics of the image, which may include, for example, the minimum and maximum luminance or the average luminance. In particular, the log average luminance is often computed to *anchor* the computation. The compression algorithm then compresses pixel contrasts according to a nonlinear function based on its luminance, as well as those global variables. No other information is used to modulate the compression curve. For instance, the key of the image can be used to determine the exponent of the gamma function [149]. In [136] and [151], an s-shaped function is defined by the image statistics, such as the mean and the variance. In [143], the histogram distribution is used to construct an image-dependent global function.

5.3.2 Linear or scaling-factor methods; lightness anchoring

The simplest TRC and the only one that preserves all image intensity ratios is a linear scaling, T(I)=s*I / $L_d(L_w) = m \cdot L_w$, where m∈[0, 1], analogous to looking at the scene through a neutral density filter [131]. The difficulty lies in finding the right scale-factor (the inverse, 1/m, is also referred to as *adaptation level* in the context of HVS). Practically, two different approaches to anchoring are known: ***a)*** the *grey world assumption* (average scene luminance is mapped to the display average, i.e., $m=L_{davg}/L_{wavg}$), and ***b)*** the *white patch assumption* (maximum non-light source luminance is mapped to the maximum displayable value, i.e., $m=L_{dmax}/L_{wmax}$).[112]

Alternatively, from a perceptual approach looking to match the impression of contrast (i.e., the visible changes in luminance) between the real and displayed image at a particular fixation point, Ward et. al proposed in [143] a method in which the scaling factor is chosen as the ratio $t(L_{da})/t(L_{wa})$, where $t(L_a)$ is the *threshold vs. intensity* (TVI) function that gives a threshold luminance that is barely visible (a.k.a *just-notable difference* or JND) for a given adaptation luminance L_a. The idea is to display bright scenes as bright and poorly lit scenes as dark, making the differences just visible in the real world just visible on the display.



a) ***b)***

Figure 5.6. Linear scaling or "hard clipping". The figures show the same synthetic HDR image linearly scaled in *a)* by $m=L_{dmax}/L_{wmax} = 10^{-3}$, and in *b)* by $m=L_{davg}/L_{wavg} =10^{-2}$.

This group of mapping methods renders acceptable results for a wide range of applications. Its strengths are its simplicity and speed, and if the right factor is chosen, the results can be acceptable for almost all applications if the raw image dynamic range is not too high. However, the process fails to preserve visibility in HDR scenes as the very bright and very dimmed values are clipped to fall within the display range.[113] Also, all images are mapped irrespective of absolute value, resulting in the loss of an overall impression of brightness [130].

---

[112] Grey world assumption has been the approach for traditional film photography, where the aperture is automatically set based on average measured light, so that it is mapped to the medium grey. Digital photography follows the white patch assumption instead, due to digital sensor linearity (which clips out-of-range luminances), which is the so-called *Expose To The Right* (ETTR).

[113] For information maximization purposes, it is desirable to choose m so that the least number of pixels are clipped. This is achieved by centring the display range in the most populated interval of world's luminances distribution.

Often, we do not have access to scene's luminances, but just to the image output of the capture process, which we assume here ideally modelled as

$$L_{sRGB}' = L_{sRGB} / 255 = (L_w / L_{w\,max})^{1/2.2} = (L_w')^{1/2.2}, \in [0, 1]$$

From now on, we will assume normalized *world* (L_w), *image* (L_{sRGB}) and *display* (L_d) values in the range [0, 1]. As we will see, most of tone reproduction operators can be expressed in terms of the captured image's luminances (L_{sRGB}) without loss of generality.

5.3.3 Non-Linear Methods

An alternative to "hard" linear clipping are non-linear TRCs, which smoothly compress the dynamic range but/hence do not preserve intensity ratios. A typical tone mapping function can be ***logarithmic***, a ***power-law*** (often referred to as a "***gamma***" function) or a ***sigmoid***, also called "***s-shape***".

5.3.3.1 Power and logarithmic

Usually, one has to choose between preserving the viewer's overall impression of brightness at the expense of the former, e.g. by using Stevens' power-law model of suprathreshold brightness and apparent contrast perception [155], as proposed by Tumblin et al. in [153]; and preserving contrast visibility, e.g. by using Weber-Fechner logarithmic relation derived from TVI functions [155], as in [132]. The election respectively results in power-law and logarithmic TRCs.[114]

$$O_{\text{gamma}}(L) = L^k \Rightarrow TRC_{\text{gamma}}(L) = (L_w^{\,k_w})^{1/k_d} = L_{sRGB}^{\,2.2 \cdot k_w / k_d},\ 0 < k < 1$$

$$O_{\text{logarithmic}}(L) = \ln(1 + k \cdot L) \Rightarrow TRC_{\text{logarithmic}}(L) \approx \frac{\ln(1 + k \cdot L_{sRGB})}{\ln(1+k)},\ k \in [0, \infty)$$

In general, when using power functions, all intensity ratios are equally compressed by a factor *k*. Logarithmic functions, instead, compress intensity ratios in bright regions, while preserving them in dark ones.

a) *power (gamma) function (k = 0.42)* ***b)*** *logarithmic (natural log., k = 2300)*

Figure 5.7. Power-law (gamma) vs. logarithmic TRCs. Despite having the same dynamic range, the perceived contrast is greater in b), while a), preserving the overall brightness, still looks somewhat washed-out.

[114] The logarithmic TRC is here provided without explicit derivation from $O_d^{-1}(O_w)$ formulation.

5.3.3.2 Rational or sigmoidal

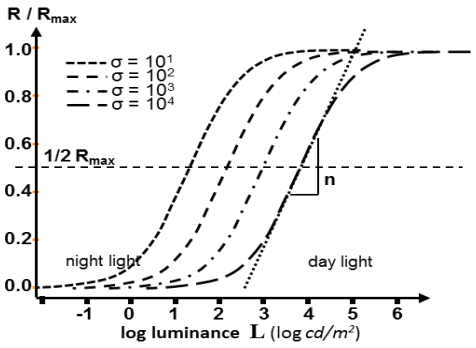
Rather than improving the perceptual performance of previous methods, some authors concentrated on improving computational efficiency and simplifying parameters, developing a family of simple and fast methods using first degree rational polynomial curves with sigmoidal shape when represented on log(L). In [151], photometric measurements of the display device are not required. Instead, only three parameters needed: highest and lowest luminance, and just noticeable difference (JND).

The use of sigmoid(al) functions has been inspired by physiological models of retinal adaptation behaviour, which can be described as the receptors' automatic adjustment to the general level of illumination, first proposed by Naka and Rushton [147] to predict, at any given adaptation level, the response of the rods and cones *R* as a function of stimulus luminance *L*; and subsequently used by other authors to psychophysically model brightness perception [149]. The model has the form of the Michelis-Menten equation: [115]

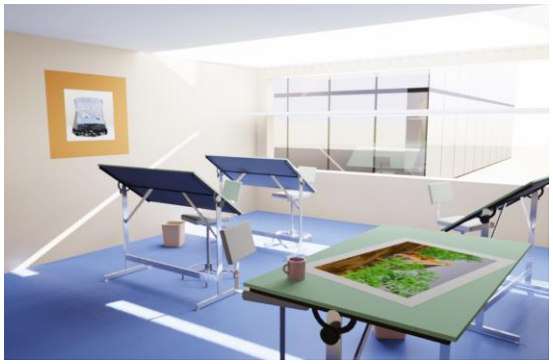
$$O_{photoreceptor}(L) = \frac{R(L)}{R_{\max}} = \frac{L_w{}^n}{L_w{}^n + \sigma_a{}^n} \Rightarrow TRC(L) = \frac{k \cdot L_{sRGB}{}^{2.2 \cdot n}}{(k-1) \cdot L_{sRGB}{}^{2.2 \cdot n} + 1} \in [0,1]$$

where σ_a, the so-called *half-saturation constant* (i.e., the level of *L* which produces half of *Rmax*), is a function of adaptation level L_a (i.e., $\sigma_a = f(L_a)$), ***R*** is the photoreceptor response, and ***n*** is computed as $(R_{max} - R_{min})/(\log L_{max} - \log L_{min})$.

The function, which aims at rendering realistic looking images in every lighting conditions, produces plausible results for many mages, since it preserves contrasts for dark image regions and asymptotically compresses image highlights so that clipping on the display can be avoided. Besides, its computation is very economic compared to logarithm or exponentiation.



a) *photoreceptor response*



b) *rational (n = 1; σ_a = 256)*

Figure 5.8. Rational or sigmoidal TRC a) Photoreceptor responses for increasing values of σ (illumination adaptation). Observe the great similarity with photographic film's characteristic curve. b) The resulting image has well balance between contrast and brightness preservation.

[115] This formula, a saturating nonlinearity, has been very useful in modeling biological behavior. It has been referred to using several names, such as *Sigmoid*, *s-shaped*, *logistic*, and with some subtle and not-so subtle derivations.

5.3.3.2.1 Holm's method

In [156][116], the image statistics determine the slope a and shift β parameters of a *sigmoidal* tone-mapping curve (or TRC), based on perceptual preference guidelines that came from the inventor's extensive experience in photographic imaging, such as the preservation of general histogram shape at the middle range by using a centering function on L* nth and (100−n)th percentile values, an L* standard deviation of around 20 and a mean L* value of around 50 (see Chapter 3, section 3.5.5). It is observed also that the L* logarithmic scale is to be preferred because it is perceptually more uniform than a linear scale and most natural images tend to have more symmetric histograms on an L* scale than on a linear scale, which makes the adjustment of the histogram spread easier (and perceptually more robust).

5.3.3.2.2 Histogram modification

The field of image processing has developed methods to adjust image contrast and visibility. One such technique is *histogram modification*, in which an *image-dependent* TRC is chosen in order to achieve a particular target luminance histogram, e.g., Gaussian or uniform. In the latter case, referred to as *histogram equalization*, the technique re-assigns luminance values to make better use of the display device range and maximise visibility and contrast by compressing the ratios (i.e., reducing contrasts and hence visibility) of underrepresented pixel intensities (i.e., pixels belonging to sparsely populated region in the scene's histogram), and vice versa [18]

$$TRC_{hist_{eq.}}(L) = \int_0^{L_{sRGB}} pdf_{L_{sRGB}}(x)dx \quad [117]$$

The method, which has the very useful feature of being idempotent [118], does well on images that have a symmetric and well-distributed histogram and, since it can dramatically improve the local visibility of details, it is very useful for medical and scientific purposes. However, it makes images look unnatural when there are large areas of dark or light background in the image (which bias the histogram toward one side), due to the exaggerated contrast and thus a harsh appearance. The issue can be lessened by specifying a particular target luminance histogram (e.g., Gaussian) instead. Although it seems unlikely that the optimal output histogram is completely independent of image content, the principle of specifying target output image tone characteristics has been incorporated into recent tone-mapping algorithms intended to improve upon histogram equalization. In these algorithms, the output histogram varies with an analysis of image content.

---

[116] The method, after [136], is part of a color reproduction pipeline created at Hewlett-Packard Labs for implementing in digital cameras an adaptive digital image tone mapping algorithm that is based on perceptual reference guidelines.

[117] Here $pdf_{LsRGB}(x)$ denotes de probability density function of the luminance levels, L_{sRGB}.

[118] An operator $T(\cdot)$ is said to be idempotent if $T(l)=T(T(l))$. This means that a compressed output image will not be further compressed if processed through the algorithm a second time. Surprisingly, not all proposed algorithms have this useful feature [131].

5.3.3.2.3 Larson's method

A modified histogram equalization method developed by Larson, et al. [143] limits the amount of gray level adjustments allowed based on luminance contrast sensitivity measurements, so that luminance differences within the image that were not visible before tone mapping are not made visible by it. Besides, the cumulative distribution function of log-luminances averaged over 1^o areas (which correspond with foveal adaptation levels for possible points in the image) is used. In [143] a model of locally adapted *glare*, *colour sensitivity*, and *acuity* is included. However, the computation is iterative, and thus the implementation is costly and slow. Thus, although images tone-mapped with the Larson's method generally have a more natural appearance than when using the traditional histogram equalization method, it does so at the cost of higher computational complexity.

5.3.4 Discussion

Global (or *spatially-uniform*) operators tend to be computationally simpler and as a result can be easier to implement and faster to perform, but present several disadvantages. In general, they fail to preserve global impression of brightness and local contrasts, which constitute the main information carrying quantity to be preserved in the final image. The simplest way to preserve contrast is to apply a linear mapping to the input image. However, linear scaling cannot map the input dynamic range to displayable values and the details in dark and light areas are clipped. While non-linear TRCs succesfully achieve dynamic range reduction, the contrast of details is compromised and the images can look washed-out.

Visual inspection of the compression curves in Figure 5.9 *a)*, may lead to the suggestion that most of these algorithms are very similar in nature, as noted in [151]. However, small differences in their functional form may lead to substantial differences in visual appearance. This is more noticeable when TRCs are plotted using log luminance scale as in Figure 5.9 *b)*. As noted before, we see that power functions, equally compress all intensity ratios by a factor *k*. Logarithmic functions, instead, compress intensity ratios in bright regions, while preserving them in dark ones. Among the presented TRCs, the sigmoidal one is the most linear one for a great portion of the input range, preserving most contrast ratios, thus resulting in the most natural images. This was previously noted by Reinhard et al. in [149]. Conversely, power-law and logarithmic TRCs have smaller slope, which translates in contrast compression for most of the input range.

Several authors have observed that, for most natural images, the distribution of log(Lw) can be well approximated by a Gaussian distribution, and thus the corresponding cumulative distribution function (i.e., the resulting image-dependent TRC) is sigmoidal. This observation connects Larson's and Holm's methods. I.e., Holm´s method can be regarded as "parametric" histogram equalization, assuming a Gaussian distribution of world luminances. However, the use of smooth and consistent sigmoidal TRC's with gentle curvature avoids the extreme contrast changes evident in some histogram equalization methods,

resulting in images that look more pleasing to the user (in terms of brightness, contrast and colour constancy) without requiring an estimation of the absolute luminance levels of the original image.

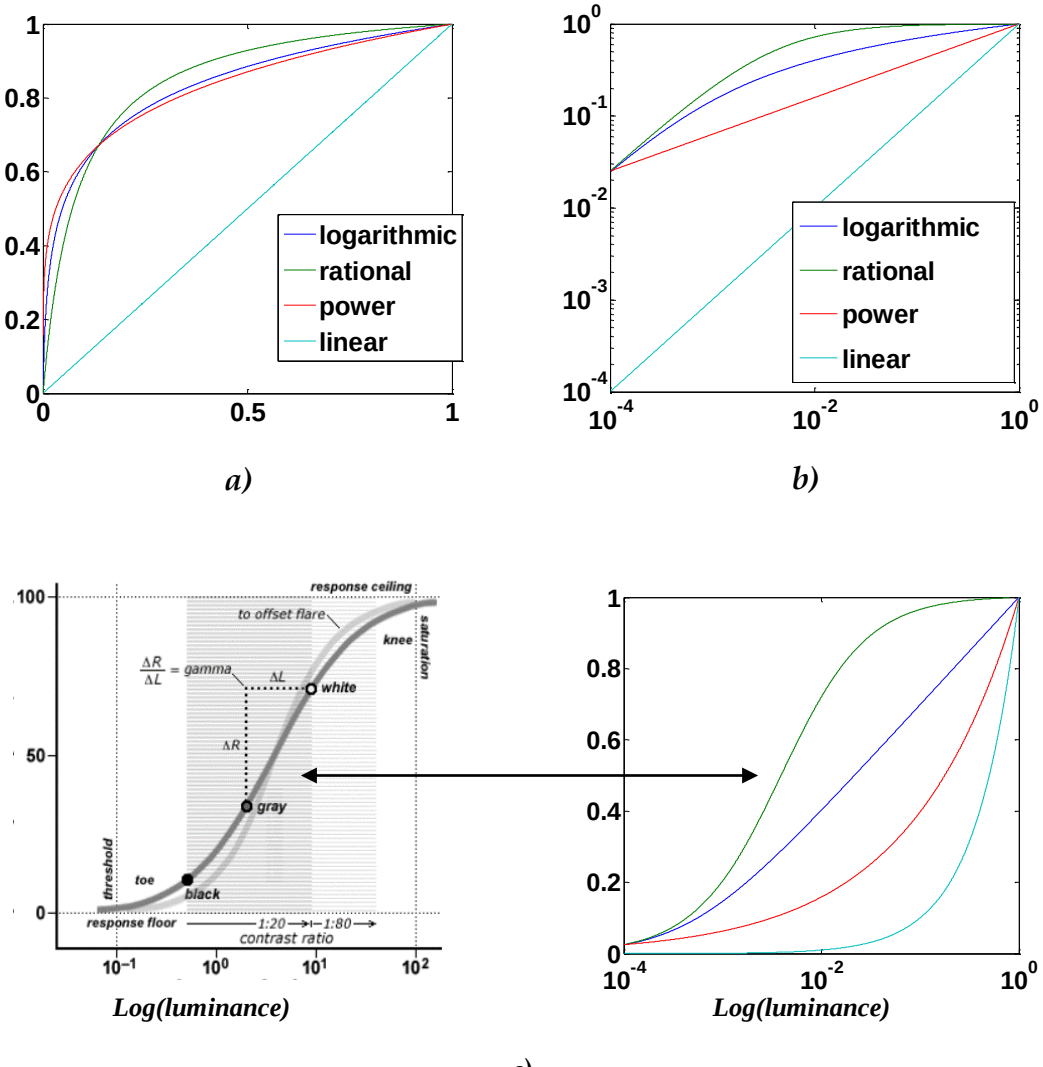


Figure 5.9. Comparison of tone reproduction curves. a) linear luminances, b) log luminances, c) similarity between the photoreceptors characteristic curve in Figure 5.3 and the sigmoidal TRC.

To compress the range while maintaining or enhancing the visibility of details, it is necessary to use more complex techniques. Further improvements can be achieved with local algorithms which can emphasize pixel visibility by considering spatial context.

Permitting the range of the two illumination regions to overlap increases the available intensity range within each region. Because TRCs need to be monotonic, or else we risk introducing reversals in local edge contrast, they cannot perform this operation. Context-sensitive algorithms (TROs) are necessary.

5.4 Spatially non-uniform techniques

5.4.1 Concept

Also called *local* tone reproduction operators, they attempt to mimic human visual system local adaptation by locally varying a mapping function parameter. The simplest formulation takes the form of a scaling factor

$$L_d(x,y) = m(x,y)L_w(x,y)$$

Algorithms that adjust pixel intensity using spatial context are commonly referred to as tone reproduction *operators* (TROs).

With local tone mapping algorithms, one input pixel value can lead to different output values depending on the pixel's surround. A local tone mapping operator is used when it is necessary to change local features in the image, such as increasing the local contrast to improve detail visibility.

These operators are permitted to transform the same pixel intensity to different display values, or different pixel intensities to the same display value. Figure 5.10 illustrates a tone reproduction operator applied to the scene image shown in Figure 5.5.

Figure 5.10. Tone reproduction operators are context-sensitive. Depending on its position in the image, a given intensity level may map into a different output level. (A) The resulting mapping is clearly not one-to-one. (B) The output image. (C) The image to display intensity mapping of the TRO is illustrated for three different spatial regions of the image. Reproduced from [131].

The guiding principle of TRO design is to preserve local intensity ratios, which may correspond to surface reflectance changes, and reduce large global differences, which may be associated with illumination differences. By following this principle, TROs are designed to match the light adaptation of the visual pathways, which discounts illuminant *extrinsic* variation but recognizes surface reflectance *intrinsic* variations.

The local luminance adaptation perceptual factor considers the fact that the eye looks at an image by scanning around. The eye can rapidly adapt to the luminance level of small regions in the original scene to enable regions in the

shadows and in the highlights to be clearly visible to the eye. In the rendered image, both the dynamic range and the adaptation environment are different. Therefore, to fully imitate the eye's adaptation process, the luminance levels of an image are adjusted according to its local luminance levels.

5.4.2 Overview of common methods

All global methods mentioned in previous section use a single adaptation value for the entire scene. Mechanistic attempts to mimic local adaptation effects by locally varying mapping function parameter(s) based on local context have been done by [141], [151], and, in a much more consistent manner, by Pattanaik et al. [148], who present the most comprehensive model of human visual system currently used in computer graphics. Ward-Larson [143] has introduced an operator that extends the work of [134] with a model of local adaptation.

In a more functional approach, more advanced local tone mapping algorithms attempt in some way or another to enhance reflectance information while compressing irrelevant illumination by implementing diverse variants of *homomorphic filtering* [119] (see [130] and [150] for a review). One may group them in the following classes:

- ***Center/surround*** methods based on the Retinex theory [142] have attempted to imitate the local luminance adaptation process by computing the new pixel values by taking the difference between the input pixel values and the average of surrounding pixels, i.e., implicit gradient estimation, in the log domain [141]. They take inspiration from the HVS receptive fields and lateral inhibition. Their common drawbacks are the creation of halos along high contrast edges and graying-out of low contrast areas.
- ***Gradient-based*** methods [133] work directly on the image gradient to increase the local contrast by weighting high and low gradient values differently, taking surrounding data into account. One difficulty of this technique is to integrate the gradient to recover the treated image.
- ***Frequency-based*** methods have also been developed, such as the bilateral filter algorithm of Durand and Dorsey [86]. The image is separated in low and high frequency bands. The low-frequency band is assumed to approximatively correspond to the illuminant and is compressed.

Observe that all these, but specially both Retinex- and frequency-based methods follow the common scheme in Figure 5.11, which is based on a multi-resolution decomposition of the image and approximate contrast in a way similar to Peli [58]: the image is first log transformed, then sent through a high-pass spatial filter which suppresses the illumination component and enhances the reflectance, and is finally exponentially transformed back into intensity space.

[119] Homomorphic filtering is a generalized technique for signal and image processing, involving a nonlinear mapping to a different domain in which linear filter techniques are applied, followed by mapping back to the original domain [18]. This concept was developed in the 1960s by Thomas Stockham, Alan V. Oppenheim, and Ronald W. Schafer at MIT [152].

The resulting displayed image has its contrast subjectively equalized so that the contrast of an object depends upon the reflectance of its background and does not depend upon whether it is fully illuminated or is in the shade. And although this technique is based on physical rather than on physiological reasoning, it is successful because it exploits the same type of enhancement which occurs in the retina.

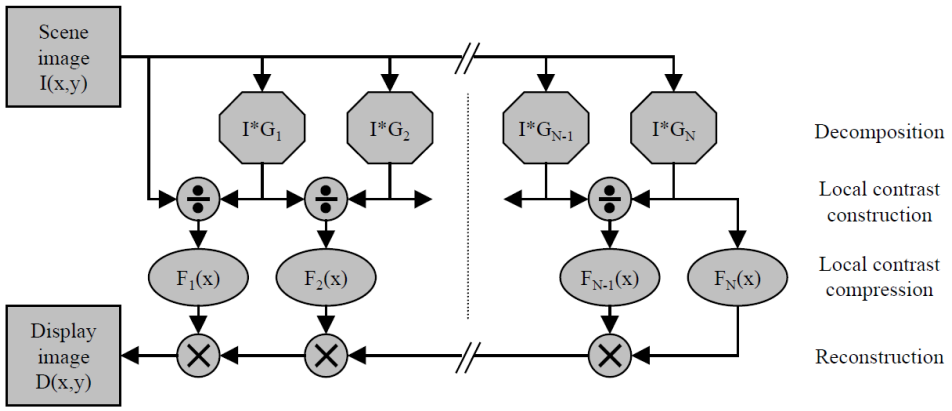


Figure 5.11.
The basic multiresolution decomposition used by TRO algorithms. G_1, G_2, G_N are (edge-preserving) smoothing filters and F_1, F_2, F_N are compression functions. Adapted from [131].

Both local adaptation and separation of reflectance and illuminance components assume local contrast variations (high spatial frequencies) correspond to surface reflectance variations, while large global variations (low spatial frequencies) correspond to ambient lighting variations. This assumption fails for sharp shadows, in the proximity of bright regions (light sources) and edges of high contrast, which are all common situations, producing haloing artefacts in the resulting images [131]. The most popular solution to overcome this is to use an edge-preserving smoothing filter to perform the frequency decomposition, thus relaxing the strong assumption that the spatial variation of the illuminant is entirely within low spatial frequencies.

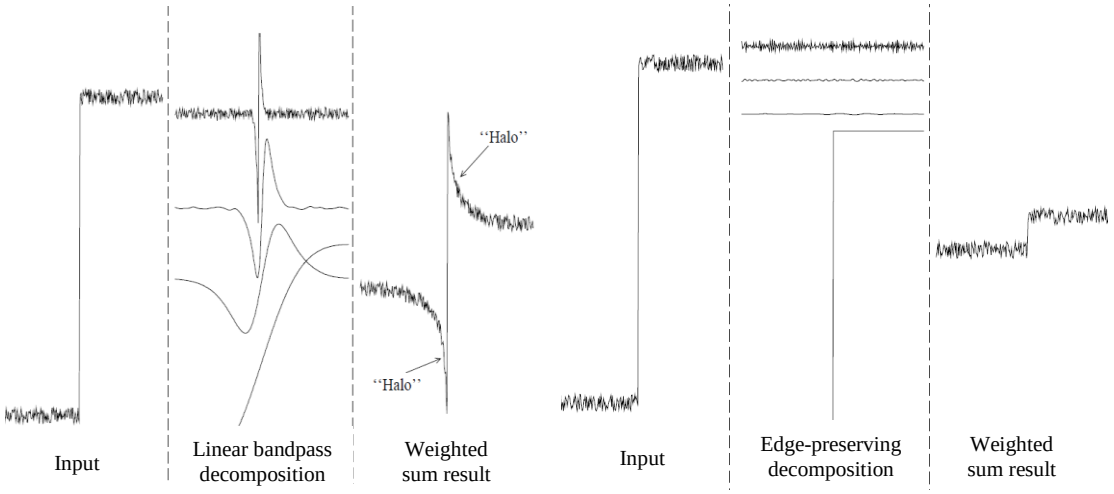


Figure 5.12. Haloing effect in a scanline from a high contrast scene. Linear filter hierarchy does not adequately separate fine details from large features. Compressing only the low-frequency components to reduce contrasts causes halo artifacts. Edge-preserving decompositions perform a much better separation. Adapted from [154].

a) Original ***b)*** Estimated reflectance ***c)*** 66% of illumination reduced

Figure 5.13. Image relighting. Given an image a) of a high-contast scene and assuming a Lambertian image formation model S = I x R, where variations in illuminance I are much larger than variations of reflectance R, we applied an edge-preserving smoothing filter on log(S) to estimate the scene's reflectance in b) as the residual. Observe that we successfully removed almost all shadows without introducing halo artefacts. Finally, a natural balanced reproduction that better resembles the scene's appearance perceived by its observers is obtained by injecting back only 33% of the removed illuminance.

5.5 Extension to colour images

Many tone reproduction operators only compress the luminance channel and apply the result to the three colour channels in such a way that the colour ratios before and after compression are preserved [151]. This is a reasonable first step, but ignores the fact that colour appearance varies with the overall intensity of the scene.

The two most common alternatives to account for colour adaptation are:

A) Linear: Von-Kries chromatic (incomplete) adaptation

$$TRC_{\text{color channel}}(L)=[D/L_a+1-D]L$$

where L_a, the adaptation level, may be determined
a) globally / locally (radius aprox. half of smallest image dimension)
b) average (gray-world) / maximum (white-patch)

The *incomplete adaptation factor D* is computed as a function of adaptation luminance L_a (20% of the adaptation white) and *surround factor F* (F=1 in an average surround).[120]

[120] In theory, this value ranges from 0 for no adaptation to 1 for complete adaptation, and in practice the minimum value will not be less than 0.65 for a dark surround and exponentially converge to 1 with increasing values of La.

B) Photoreceptor-like adaptation:

This is the sigmoidal TRC presented in Section 5.3.3.2, but applied to each color channel independently.

Recently, some authors have shown with succes how to solve the original tone reproduction optimization problem (Figure 5.2) by appling complex image appearance models to the scene and the display observers. While this represents the state-of-the-art perceptual tone reproduction, it requires complete characterization of scene and display luminances , which we lack. Besides, its is not easily generalizable to tone reproduction intents other than perceptual one.

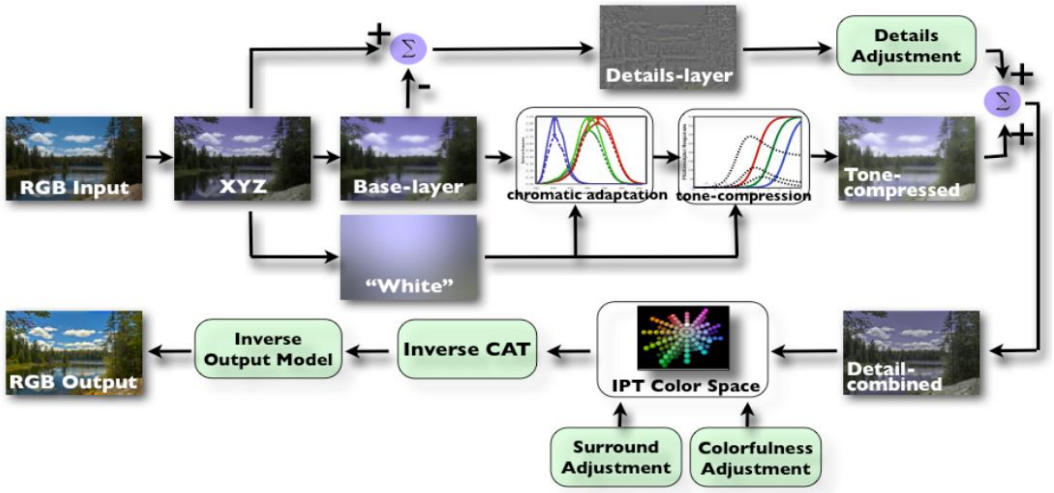
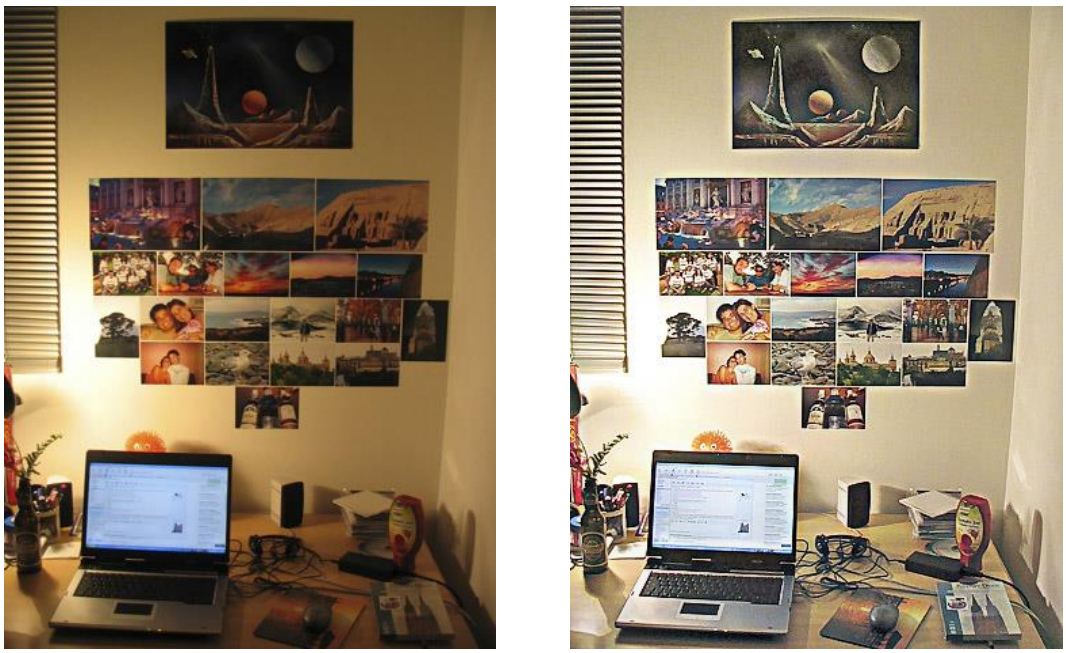


Figure 5.14. Flowchart of iCAM06 image appearance model applied to the perceptual reproduction of high dynamic range (HDR) scenes. Reproduced from [50].



a) original image ***b)*** corrected image

Figure 5.15. Correction of image with different illumination colours. A local Von-Kries chromatic incomplete adaptation mechanism allowed us to reproduce the visual appearance of this scene with mixed illumination colours. We also enhanced local contrast to increase detail visibility.

5.6 Summary

Appropriate tone reproduction is needed in diverse applications, from artistic photography to scientific imaging, to produce a visual match, in terms of *naturalness* and *usefulness*, between the scene's and image's hypothetical observers to the limits possible on a given display media. Particularly, it is noticed that colour images differ from direct human viewing by the lack of dynamic range compression and colour constancy. It is therefore our aim to present a generally applicable operator that is fast and practical to use and provides intuitive user parameters.

Inspired by early vision mechanisms, we review the work to date on tone reproduction techniques that mimic some characteristics of the human visual system, in particular, colour and lightness[121] constancy, to discount the illuminant or, equivalently, relight the image. While we use visual models to relate the perceptual responses of a hypothetical scene observer to the responses of the display observer, thus providing a theoretical basis for perceptual tone reproduction, note that we do not attempt to build a full visual model which may account for many perceptual phenomena but will most likely be too complex for our purposes. Rather than contributing with new models to the wealth of emerging appearance models, our main contribution here is to expose the problem, providing qualitative research and appropriate tools for early processing. We used findings from sensory adaptation mechanisms (such as photoreceptor gain control) and cognitive mechanisms (such as perceptual constancy[122]) to motivate the design and understanding of algorithms but made engineering-based design decisions where appropriate. In particular, no claim is made that HVS behaves the way we describe.

We have shown by example that good tone reproduction does not require ad-hoc methods and subjective judgements. We have reviewed brightness, contrast sensitivity and photoreceptor -based global tone mapping curves, that result from chaining scene's and display observer models, and that are good enough in terms of local contrast (detail) visibility and global brightness appearance, as the two most important information-carrying (hence preferential to image quality) attributes to be preserved or enhanced, (for the sake of naturalness or usefulness, respectively) in the final image.

In addition to this functional separation, we have further outlined how to mimic also human visual system local adaptation by using edge-preserving image smoothing to locally vary a mapping function parameter. From high-level computational approach, we have seen the connection to intrinsic image models where the hypothesized functionality of the Human Visual System is the

[121] Lightness is defined as the perceived reflectance of a surface. It represents the visual system's attempt to extract reflectance based on the luminances in the scene.

[122] The ability of a vision system to discount the accidental conditions (e.g., the colour of the illuminant) and to extract (scene's) invariants (e.g., object reflectance) is referred to as *perceptual constancy*.

recovery of scene's intrinsic properties, specifically separation of *reflectance*, which actually conveys useful information, from *shading* (variations in illumination), which we regard in the context of image improvement as unwanted distortion to be removed. This poses the problem in terms of: 1) spatial decomposition of the intensity signal and 2) spectral normalization of the surface reflectance and the effective compression of the dynamic range.

We notice also that the presented methods have a similar mathematical structure and therefore can be considered virtually equivalent to histogram equalization for efficient representation and homomorphic filtering to invert the image formation process, which resembles image restoration.

It seems, however, that there is no a single tone reproduction method that works well for all scenes. The development of such a method seems to be at present unattainable given the current status of computational models of human visual response. The lack of comprehensive image metrics in graphics also limits the study. At present, the answer is to select the most appropriate method for a given task and adjust its parameters to get best possible results. Depending on requirements, a number of different operators are available for use and they must be selected on the premise of the 'best tool for the job'.

Future work should include a review of very recent research in [125], [127] and [135], as well as an extensive validation of tone reproduction operators, preferably through psychophysical comparison.

REFERENCES

[124] ADELSON, E. H. Lightness Perception and Lightness Illusions. *The New Cognitive Neurosciences,* 2nd ed., M. Gazzaniga, ed. Cambridge, MA: MIT Press, pp. 339-351, 2000.

[125] BARRON, Jonathan T.; MALIK, Jitendra. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 2015, vol. 37, no 8, p. 1670-1687.

[126] BARROW, H.G. and Tenenbaum., J.M. *Recovering intrinsic scene characteristics from images*. Technical Note 157, AI Center, SRI International, April 1978.

[127] BI, Sai; HAN, Xiaoguang; YU, Yizhou. An L1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. *ACM Transactions on Graphics (TOG)*, 2015, vol. 34, no 4, p. 78.

[128] BRAINARD, D.H. and Wandell., B.A. Analysis of the retinex theory of colour vision. *J. Opt. Soc. Am. A*, Vol. 3, No 10, October 1986.

[129] ČADÍK, Martin, et al. Image attributes and quality for evaluation of tone mapping operators. *National Taiwan University*. 2006.

[130] DEVLIN, Kate. A review of tone reproduction techniques. *Computer Science, University of Bristol, Tech. Rep. CSTR-02-005*, 2002.

[131] DICARLO, J.; WANDELL, B. Rendering high dynamic range images. *Proceedings of the SPIE: Image sensors*. 2000. p. 392-401.

[132] DRAGO, Frédéric, et al. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*. Blackwell Publishing, Inc, 2003. p. 419-426.

[133] FATTAL, Raanan; LISCHINSKI, Dani; WERMAN, Michael. Gradient domain high dynamic range compression. *ACM Transactions on Graphics (TOG)*. ACM, 2002. p. 249-256

[134] FERWERDA, James A., et al. A model of visual adaptation for realistic image synthesis. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996. p. 249-258

[135] GU, Bo, et al. Local edge-preserving multiscale decomposition for high dynamic range image tone mapping. *IEEE Transactions on image Processing*, 2013, vol. 22, no 1, p. 70-79.

[136] HOLM, J., "Photographic Tone and Colour Reproduction Goals," CIE Expert Symposium on Colour Standards for Image Technology, pp. 51–56 (1996);

[137] HORN, B.K.P. Determining lightness from an image. *CGIP 3*, 4, pp. 277-299, December 1974.

[138] HUNT, R.W.G. Tone Reproduction, in The Reproduction of Colour, 6th Edition, John Wiley & Sons, Ltd, Chichester, UK. 2004.

[139] HURLBERT, A.C. The Computation of Colour.

[140] HURVICH, L.M. and JAMESON, D. An opponent-Process Theory of Colour Vision. In Yantis, S. ed. *Visual perception: Essential Readings*. Taylor & Francis Group, 2001(?)

[141] JOBSON, D.J., RAHMAN, Z. and WOODELL, G.A. A Multiscale Retinex for Bridging the Gap Between Colour Images and the Human Observation of Scenes. *IEEE Trans. on Image Processing*, Vol. 6, No. 7, July 1997.

[142] LAND, E.H. Recent advances in Retinex theory and some implications for cortical computations: colour vision and the natural image. *Proc. Nat. Acad. Sci. USA* 80, pp. 5163-5169, 1983.

[143] LARSON, G.W., RUSHMEIER, H. and PIATKO, C. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 1997, vol. 3, no 4, p. 291-306.

[144] MACEVOY, B. *Color Vision: adaptation, anchoring & contrast*. Online resource at https://www.handprint.com/HP/WCL/color4.html.

[145] MANTIUK, R. et al. High dynamic range imaging pipeline: perception-motivated representation of visual content. *Human Vision and Electronic Imaging*. 2007. p. 649212.

[146] MORONEY, N. Local Colour correction using Non-Linear Masking. *IS&T/SID Eighth Colour Imaging Conference*.

[147] NAKA, K. I.; RUSHTON, W. A. H. S-potentials from colour units in the retina of fish (Cyprinidae). *The Journal of physiology*, 1966, vol. 185, no 3, p. 536-555.

[148] PATTANAIK, Sumanta N., et al. A multiscale model of adaptation and spatial vision for realistic image display. En *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM, 1998. p. 287-298.

[149] REINHARD, Erik, et al. Photographic tone reproduction for digital images. *ACM transactions on graphics (TOG)*, 2002, vol. 21, no 3, p. 267-276.

[150] REINHARD, Erik, et al. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[151] SCHLICK, Christophe. Quantization techniques for visualization of high dynamic range pictures. *Photorealistic Rendering Techniques*. Springer, Berlin, Heidelberg, 1995. p. 7-20.

[152] STOCKHAM, T.J., Jr. "Image processing in the context of a visual model". *Proc. IEEE 60*, 7, July 1972, 828-842.

[153] TUMBLIN, Jack; RUSHMEIER, Holly. Tone reproduction for realistic images. *IEEE Computer graphics and Applications*, 1993, vol. 13, no 6, p. 42-48.

[154] TUMBLIN, Jack; TURK, Greg. LCIS: A boundary hierarchy for detail-preserving contrast reduction. *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999. p. 83-90.

[155] WANDELL, B.A. *Foundations of Vision*. Stanford University. Sinauer Associates, 1995.

[156] ZHANG, Xuemei, et al. System and method for digital image tone mapping using an adaptive sigmoidal function based on perceptual preference guidelines. U.S. Patent No 7,023,580, 4 Abr. 2006

[157] *Perceptual and Artistic Principles for Effective Computer Depiction* Course Notes for SIGGRAPH 2002, San Antonio, Texas. July 2002.

Chapter 6

KNOWLEDGE-BASED IP SYSTEM IMPLEMENTATION

It could be easily concluded from previous chapters that image processing is a subject that lends itself to a rigorous, mathematic treatment and, thus, it is often perceived as being rather theoretical. Nevertheless, Image Processing (IP) has seen applications in numerous fields such as medicine or astronomy, what requires that complex IP is made accessible to the *non-expert* user.

For image processing, algorithm development environments, function libraries, source code repositories, and specialized IP tools and packages such as Photoshop or MATLAB Image Processing Toolbox [173], become extremely useful for rapid prototyping of algorithms. However, all these provide just a syntactical integration of IP programs. If there is no need to understand the underlying algorithms, but there is a need to perform specialized or complex IP requiring IP *expertise*, then a *knowledge-based* application for IP is needed.

While previous chapters dealt with simple, low-level, IP *routines* such as noise removal or tone mapping, this chapter deals with complex, high-level, IP *tasks*, where the operators and the order in which they are used is not fixed, but depends on the final goal, the context and the image content. Tools provided today by image processing libraries can become highly technical and non-intuitive including various gauges and knobs. The application of IP techniques

to real world problems, such image quality improvement, still presents two major issues.
First, most of IP application systems are developed as special purpose ones that work only for the given conditions.

Second, while IP has become highly specialized with programs implementing more and more complex functionalities, no support is provided to inexperienced users, who just have a basic understanding of the field of image processing and its terminology, to solve practical image processing problems. Difficulties frequently arise when it comes to find out, set the value of parameters and form a sequence of classical IP operators that meet the requirements of specific user tasks. Moreover, such a process typically involves empirical knowledge and thus it is not suitable for routine application.

The design of knowledge-based systems exclusively dedicated to IP, i.e. those that automate the performing of IP tasks restricted to image transformations without any interpretation of the image content, is an important issue because it opens perspectives in two major trends: *a)* to make IP accessible to end-users while limiting their cognitive and skill requirements; and *b)* to improve the performance of vision systems and interpretation systems. Such systems have a capability for self-configuration to different IP requests and application contexts by using explicit knowledge representation about image processing techniques. They provide a solution where algorithmic ones either do not exist or are very costly to implement.

Our approach to this problem belongs to the more general category of *program supervision* systems, where IP knowledge from an expert not necessary aware of algorithm implementation is used (usually in the form of rules) for dynamic composition of image processing through the *selection*, *parameter tuning* and *scheduling/execution* of existing operators from a library to accomplish a user's processing objective, while at the same time keeping some user interaction regarding control and influence in the decisions [166]. Such a system serves also as a prototyping environment in which existing functions can be integrated with new code to gain flexibility and reduce development time.

The present chapter describes the development from scratch of such a system, adjusted to the actual needs of the thesis. This includes, but is not limited to, recognizing what knowledge is being used to solve the problem, categorizing it and determining the best way to represent it. As image quality improvement is a general problem arising in various application domains, we are interested in providing both *knowledge models* and *software tools* which are independent of any particular application and of any IP package, i.e. they should be as general and flexible as possible. But we are not only interested in solving IP problems, we also want to understand the reasoning of the system, in order to improve its results and behavior, and, in the long term, to allow explaining and teaching Image Processing to non-expert users.

6.1 Framework: Problem solving in Image Processing

IP has become highly specialized with programs implementing more and more complex functionalities, and despite the fact that every end-user cannot have a deep understanding of program semantics and syntax, programs are most of the times integrated (in the library) just from a low-level point of view [170].

6.1.1 Distinctive features of image processing

IP is achieved by means of *operators* (also called *subprograms*, *primitives*, *routines* or *procedures*). An operator is a program characterized by its inputs, parameters and outputs. The inputs consist in images which can be either purely digital or more symbolic. Parameters are necessary to adapt the behavior of the operator with regard to the specificities of input images and the objectives to be reached. The outputs can take the form of digital or symbolic images, as well as digital or symbolic attributes. There exists a great variety of operators in the literature; some of them can physically modify pixel values (smoothing, thresholding...), others ensure the construction of a new representation of the image data (regions, adjacency graph of regions, quad-tree...), others can also calculate the value of attributes (texture, number of regions...). Solving an IP problem consists in selecting operators, finding the optimal values for their parameters, and organizing operators into suitable sequences.

6.1.2 Integration of image processing operators

In order to facilitate the wider use of digital image processing techniques, three main categories of image processing software have been developed in the last decades: ***packages***, ***interactive systems*** and ***languages***, which respectively address the need for *transportability*, *improved usability* and *performance power*. These have already been described in an overview of image processing software contained in [171]. From now on this chapter concentrates on the use of packages.

Many image processing packages consisting of an image processing library and a simple test or programming environment have been developed in the last decades, where the individual programs are integrated from a low-level point of view [170]. Good examples are Khoros[123] and ImageJ[124]. These packages provide however little help to the user without enough expertise for digital image

[123] Khoros Pro 2001 of Khoral Inc. [172] is an integrated development environment for IP with a special module for teaching known as the "Digital Image Processing Course". Khoros has earned its place as a pedagogical platform for IP mainly because it offers a visual programming environment coupled with an easy way to link C functions. It also has a large base of users who are willing to exchange their knowledge.

[124] ImageJ [175] is a public domain, Java-based, powerful, full-featured image processing program developed at the National Institutes of Health. ImageJ was designed with an open architecture that provides extensibility via Java plugins. Custom acquisition, analysis and processing plugins can be developed using ImageJ's built in editor and Java compiler. User-written plugins make it possible to solve almost any image processing or analysis problem, what has made ImageJ a popular platform for teaching image processing.

processing in order to solve practical problems such as image quality improvement. Difficulties arise however when it comes to precisely defining a request, working out a solution and evaluating the results [162]. The following are popular problems encountered when solving real-world IP problems [158], [159], [165]:

- ***Defining a request.*** Formulating the corresponding IP request to a problem set by an expert from the domain of application (biology, geography…) is not straightforward and requires selecting relevant tasks and judicious quantitative and qualitative image features [158].
- ***Working out a solution.*** Because there is no general standard of image quality, most of the enhancement techniques in existence to date are empirical or heuristic methods, dependent on the particular type of image [8]. More important, most of these techniques require interactive procedures to obtain satisfactory results, and therefore are not suitable for routine application. In general, according to *i)* image content and *ii)* processing purpose, effectively realizing complex IP tasks requires:
  - **Selection of appropriate operators.** At present, the answer is to use the most appropriate method for the situation. Depending on requirements, a number of different operators are available for use and they must be selected on the premise of the 'best tool for the job'.
  - **Determination of optimal parameters**: Besides requiring the user interaction, complex image processing tasks require setting the correct parameters values and, therefore, are often difficult to fine-tune.
  - **Combination of primitive operators**: it is often necessary to combine many primitive operators to perform a meaningful task. For example, a popular way of extracting regions from an image is to apply edge detection -> edge linking -> closed boundary detection.
  - **Execution (trial-and-error experiments)**: because IP expertise is essentially intuitive and it is very hard to estimate a priori the performance of an operator for a given image, IP experts often proceed by a trial-and-error process.
- ***Evaluating the results.*** IP is a domain where no ideal evaluation function exists and no one but the domain expert can perform the final validation of an application. This validation can either be done visually, or be based on a more global testing protocol (e.g. statistics on object features extracted from the image) [8]. In Chapter 3 objective methods for image quality assessment were introduced.

In order to overcome such difficulties, the IP expert has to make the most of the know-how acquired when developing previous applications. In this chapter, we are interested in developing software tools to help non-specialists in IP to work out a solution to a given IP request.

6.1.3 Knowledge-based Integrated IP Systems

Various intelligent methodologies generally based on artificial intelligence techniques have been investigated to develop intelligent IP systems. According to the nature of the knowledge these systems contain explicitly, they have been classified in [158] as:

- ***intelligent algorithms:*** express in detail how programs work, and describe explicitly the internal mechanisms of the algorithms.
- ***intelligent interpretation systems:*** contain explicit knowledge on the modelling of objects of the world. They are also referred to as *image understanding systems* (IUS's) [165].
- ***intelligent integrated systems:*** contain all the knowledge needed for the selection and the use of programs seen as black boxes. In such intelligent integrated systems, what programs do (their goal) and when (under which conditions) are expressed explicitly. They are also referred to as *expert systems for image processing* (ESIP's) [165].

In this thesis, we concentrate on the last ones, i.e. knowledge-based systems exclusively dedicated to IP, whose major purpose is to automate the performing of IP tasks (referring to general IP objectives such as segmentation, restoration or enhancement of images) restricted to image transformations without any interpretation of image content. Users describe tasks to perform on images and the system constructs a specific plan, which, after being executed, should yield the desired results.

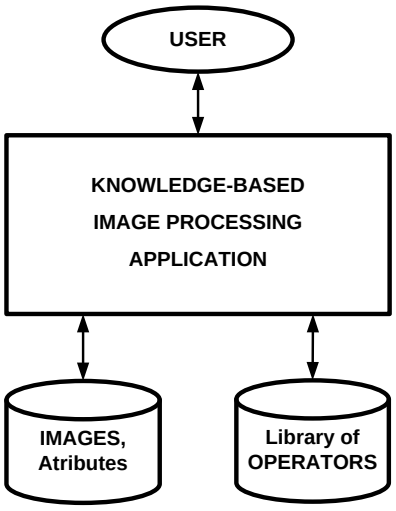


Figure 6.1. Use of a knowledge-based application for image processing.

The design of such systems, which continues to present a major challenge, is an important issue because it opens perspectives in two major trends [158]:

- To make image processing accessible to non-specialists in image processing such as biologists or astronomers, by means of programming environments that can cover a wide range of tasks and contexts, while limiting cognitive and skill requirements of users, from an advisory guide up to fully automatic program monitoring systems.
- To improve performances of autonomous vision and interpretation systems working in complex and variable environments, where a well known reason of failure is the weakness of the IP level.

6.1.3.1 Knowledge-rich paradigm: program specification by abstract command

While it has been argued that this reasoning mechanisms should be based on knowledge represented dynamically, most of recent proposed solutions non-dedicated to any specific application are all based on the search of processing schemes adapted to the nature of the problem and the images, among a base of predefined image procedures [160].Instead, knowledge-rich problem-solving relies on the existence of an expertise modeled in the knowledge base describing an abstract plan adapted to a given task. This expertise includes knowledge about conditions of applicability and knowledge about behavioral adjustment to particularities of the context [159].

Approaches belonging to the more general category of program supervision systems consider the problem as the dynamic building of chains of IP through the selection, parameter tuning and scheduling of existing operators [166]. In order to find out, set the value of parameters and form a sequence of IP operators according to the specific features of the problem and the characteristics of the image, it is necessary that such systems have capability for self-configuration to different IP requests and application contexts by the way of reasoning based on an explicit and operative modeling of the knowledge that image processing researchers have acquired and accumulated through the development of image processing techniques [160].

In this context, the terms 'scheme' or 'workflow' refer here to a hierarchical plan coded with production rules and frames of abstract modules, each one corresponding to simple IP tasks such as noise removal, contrast enhancement or edge detection.[125] The executable procedures are actually built by instantiating each module, and may be controlled interactively by the user's feedback [162].

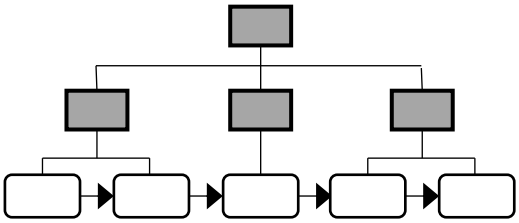


Figure 6.2. Plan Representation

Although we usually have to improve the composed program in its analysis capability and efficiency, the systems of this type facilitate the use of IP packages in order to perform complex IP tasks, since the user only has to specify the composition of programs in the library without worrying about their detailed syntactic and semantic structures, such as argument data types. The system stores that information and, based on it, determines real arguments for each subroutine, if necessary, asks the user to specify missing parameters, and generates a complete sequence of programs calls.

[125] In image processing hierarchy is natural. The processing of an image is not a monotonic process, for which each operator makes intermediate data progress towards the solution in a continuous and uniform way. Instead, one generally has to consider several intermediate steps which are not directly related to the final goal (e.g., contrast enhancement, noise reduction, pre-segmentation) [159].

6.2 Solution description

This section is devoted to describing the design of the system we have implemented using knowledge-based techniques to automate execution of IP, i.e., given an image to be processed, a goal to reach (e.g., enhance its quality), and constraints on the result (e.g., naturalness), the system generates and performs the processing.

The system is conceived as an initial prototype. During its development, we are mainly concerned with the problems of modeling both the knowledge and reasoning used in IP, being the system's flexibility and generality the only strict requirements we have. For their completeness, soft requirements are implicitly outlined by first presenting the motivations and objectives that lead us to the adopted architecture. Then, the knowledge representation paradigms chosen to model the several types of knowledge involved are described. Finally, this section concludes with a description of the type of reasoning and techniques used to find a solution.

6.2.1 Motivations and Objectives

Following the ideas derived from the study of knowledge-based systems for IP described in the previous section, our motivations for building a new *intelligent* IP system stem from two major objectives: ***a)*** providing an experimentation tool for IP and ***b)*** aiding inexperienced users in application fields manage image processing techniques that are needed for their applications. These are detailed in what follows.

6.2.1.1 Experimentation tool

While development of algorithms is usually based on theoretical frameworks, development of image enhancement techniques almost always requires extensive experimental work with large sets of sample images involving algorithm composition, execution, revision and comparison of candidate solutions. An explicit knowledge representation together with execution automation capabilities may greatly improve such highly empirical or heuristic tasks. This would also allow an efficient communication of the resulting IP expertise and rapid production of results, favoring dialogue and cooperation between both, domain and IP experts.

In this respect, the here proposed system can be seen as an experimentation tool for interactive algorithm development and comparison of IP techniques, where algorithms can be applied to images, their attributes may be modified and the interrelation between results studied.

6.2.1.2 IP Support tool

While different application areas such as medicine, astronomy or photography may share similar IP needs (e.g., enhance the quality of captured images), depending on each case they frequently require applying specific complex

techniques adapted to the characteristics of each one. By encapsulating the knowledge of program use in a knowledge data base and by emulating the strategy of an IP expert in the use of the programs, we could support non-specialists in IP by providing them with a knowledge-based system for IP, which enables both ***i)*** developing their own IP applications while limiting their cognitive and skill requirements, and ***ii)*** exchanging knowledge bases for different application purposes.

Since image quality improvement is a general problem arising in various application domains, we are interested in separating the description of IP procedures from actual algorithm implementation, providing both knowledge models and software tools which are ***general*** (independent of any particular application and of any IP package) and ***flexible***.

Last, but not least, we are not only interested in solving IP problems, we also want to understand the reasoning of the system, in order to improve its results and behavior, and, in the long term, to allow explaining and teaching Image Processing to non-expert users.

6.2.2 System Overview

We develop our system from scratch, based on a mixture of image enhancement heuristics and procedures taken from image processing and photography communities, that employs production rules as the main knowledge representation paradigm not only to describe IP expertise as an ordered sequence of abstract commands, but also to explicitly state the relation between images, their attributes and the IP programs in the library.

Small systems like this one, developed by the expert himself, can be very helpful during software development for simulation and prototyping, providing a flexible way to encode and modify the knowledge base over time as they are discovered are ideal for training new knowledge engineers.

Our system is based on a typical production system architecture, with a global database of facts or assertions about the problem, a set of rules which constitute the program, stored in a rule memory of production memory, and an inference engine, required to execute the rules. As such, it necessarily resembles both general purpose knowledge-based shells, such as Personal Consultant[126], and expert systems for image processing, such as EXPLAIN [168].

The following sections describe in detail the system architecture, the modeling of IP domain knowledge and the inference engine.

---

[126] Personal Consultant was developed by Texas Instruments (TI) of Austin, Texas, using Scheme, a variation of LISP. It is a backward-chaining, rule-based system based on the classic EMYCIN that supports limited forward chaining and uses an inference network to represent its knowledge base internally.

6.2.3 System Architecture

The system consists of four major components: the ***graphical user interface*** (GUI), the ***IP modeling subsystem***, the ***execution subsystem***, and the ***knowledge management subsystem***. In the following each one is described in detail.

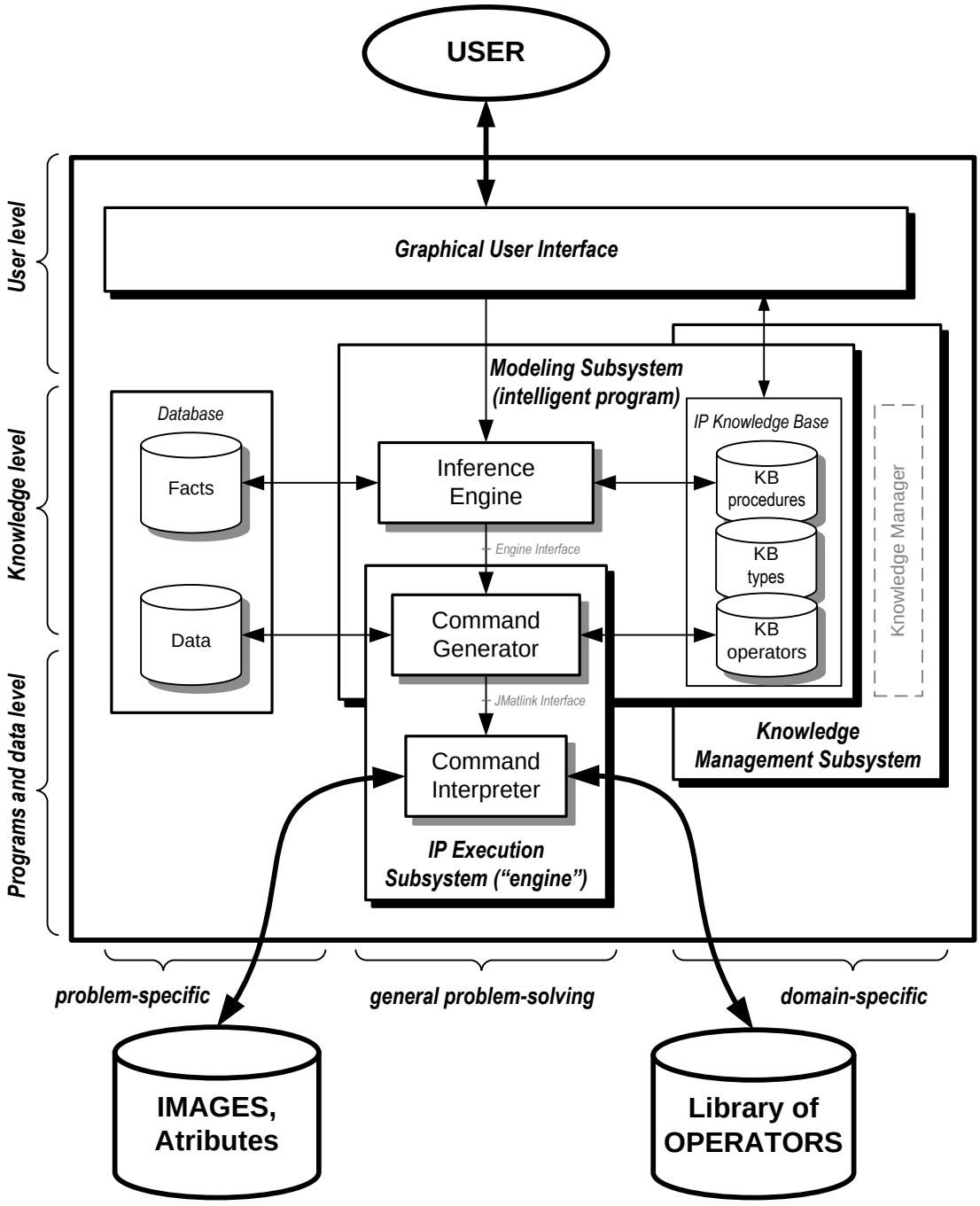


Figure 6.3. System architecture. The figure shows the four overlapping modules that compose the internal architecture of our system application. It makes explicit both the vertical division in *programs* and *data*, *knowledge* and *user* abstraction levels; and the horizontal division in *problem-specific*, *general problem-solving* and *domain-specific* aspects. The relation to external components is also shown. This architecture relies on a data store, or *working memory*, serving as a global database of symbols representing facts or assertions about the problem; on a set of rules which constitute the program, stored in a *rule memory* of production memory; and on an *inference engine*, required to execute the rules.

6.2.3.1 Graphical User Interface.

The user interface serves to provide the end user with a friendly means of communicating with the intelligent program. It does this by providing convenient interactions using menus, natural language, and/or graphical displays. This interface can be used for several purposes, such as enabling the intelligent program to pose questions to the user about the problem at hand; providing explanations about why or how these questions and/or decisions where taken; displaying the derived results; providing graphic output for the derived results; allowing the user to save or print the results; etc.

6.2.3.2 Modeling Subsystem.

This is namely the intelligent program, which interacts with the user and aids him to perform an IP task. It is divided in two components, a ***knowledge base*** and an ***inference engine***, showing a clear separation of the knowledge from its use:

a) ***knowledge base***: contains all of the relevant, domain-specific, problem-solving knowledge about IP procedures, algorithms and image attributes, that has been gathered by the knowledge engineer.

b) ***inference engine***: contains the general problem-solving knowledge required to infer an appropriate ordered set of symbolic actions for a given user request, according to the contents of the knowledge base and the data accumulated about the current problem.

Closely associated with the modeling subsystem is a ***data*** or ***fact base***, which contains the problem-specific data, i.e. the initial information provided by the user about the current problem as well as information progressively derived as it is solved. Here, the term information refers to both, images and additional data related to them, such as measurements.

6.2.3.3 Executing Subsystem

Also called ***engine***, it first uses the syntactic information about the IP algorithms to convert each of the symbolic actions chosen by the modeling subsystem into an actual procedure call. In fact, this component acts as an interface between the modeling subsystem and the actual IP package component (here referred to as the ***core*** of the engine). The latter consists of an IP library and a programming environment, and performs the actual processing. Having such an interface between both subsystems allows a great architectural flexibility: they do not need to be in the same machine, or implemented using the same language. Moreover, depending on the specific purpose, different IP components may be interchanged or used simultaneously with the same modeling component.

6.2.3.4 Knowledge Management subsystem

This is the part of the system in charge of maintaining the knowledge bases, what includes but is not limited to tasks such as parsing the knowledge bases from and to a readable format. Its development remains elemental and needs to be improved in the future.

6.2.4 Modeling of IP domain knowledge

First, knowledge models related to the activity of an IP expert processing images using an IP library have to be derived. This is a knowledge-engineering task involving recognizing what knowledge is being used to solve the problem, its categorization, and determining the best way to represent it.

Since IP problem solving consists in arranging relevant operators to achieve a processing goal, a knowledge base for IP should describe the operators with their arguments, the conditions under which they are applicable, the possible relations between them, etc. This knowledge about the operators of the library should be complemented with *i)* knowledge about the domain of IP and its context application (to describe a priori information about image domain or expected results), *ii)* knowledge about the expertise in IP (i.e., applicable techniques and procedures according to image content), and *iii)* knowledge about the control of the resolution [160].

Some of this knowledge consists simply of facts, while other identifies relationships between them. Some knowledge is algorithmic while other is heuristic. Some knowledge is declarative, while other is procedural.

There are two kinds of knowledge handled by the system: ***i)*** the knowledge independent of the contents of a given image; and ***ii)*** the knowledge dependent on them. The former is the static or descriptive knowledge of image data types and image processing algorithms, while the latter is the heuristic criteria of image processing procedures based on the experiences of experts on image processing. This knowledge should be described independently of any application domain, of any program library, and of the implementation language of the knowledge-based system (in our case Java) by using a language with a formalized syntax, well-defined semantics and expressivity capacity.

We take a hybrid approach, in which the knowledge description language (which in an initial prototype is used as a common human readable format for writing, consulting, and exchanging knowledge bases) formalizes previous knowledge using two types of declarative descriptions: structural frame-like and rule-oriented ones. Structural descriptions are used for data types and algorithm description, while IF-THEN type rules are used for expression of heuristic criteria and image processing procedures.[127]

---

[127] Five major knowledge representation schemes are commonly used in knowledge-based systems: *logic*, *rules*, *associative (i.e. semantic) networks*, *frames* and *objects*. Each one uses different types of reasoning techniques to interpret and apply the represented knowledge. Many hybrid systems use frames to represent structured knowledge and rules to reason about the former. Another use of frames is the representation of factual knowledge within a rule-based system. All facts (whether initial facts, intermediate results, or final conclusions) can be implemented within frames using slots like: *value,…*; *certain factor,…*; *possible-values,…* and so forth. This use of frames has occurred in the implementation of inference systems [163], [164].

6.2.4.1 Knowledge independent of the contents of an image

This is the basic knowledge of the different data types and the function and usage of each image processing algorithm. These basic entities of knowledge can be described in a static way using a structure with a set of attribute-value (often called *slot-fillers*) pairs. This allows a hierarchical representation using inheritance both to reduce the amount of knowledge description and to introduce flexibility to the system [169].

6.2.4.1.1 Operators or Algorithms

IP tasks can take two forms: either a decomposition into sub-tasks, or a call to a particular program module in the library [162]. In an initial phase, we only consider the latter. From a user's point of view, an IP program can be seen as a black box: all that is known about it is the nature of transformations performed on inputs to produce outputs and the only possible action is the tuning of parameters. Description of its usage, such as a list of arguments data, preconditions and/or effects, the way to run it, etc., is usually documented in the manual of the library. Such structured information about program modules can be very useful as the knowledge source for a structural algorithm representation, as is done in consultation systems for IP, where it is frequently used to help a user to select an appropriate command and its parameters.

In the implemented system, the term ***algorithm*** is used to denote such a user-oriented representation of each program in conceptual terms (goal, inputs, parameters, outputs, calling syntax, performance, resources, etc.) with a link to the code enabling to run it. Each algorithm, which is defined independently from any specific implementation, is described by the following fields:

- Name, which provides a functional description.
- Arguments' (data and parameters) names and types.
- Description of the calling syntax.
- Comment (a brief explanation of the operator).

```
Algorithm brighten {
·Matlab  Name: brighten
·Input  Names: image,amount
·Input  Types: Image,percentage
·Output Names: brighten image
·Output Types: Image
·Comment: this algorithm is used to
brighten  an  image  some  amount
between 0 and 100%
}
```

Figure 6.4. Example of *algorithm* definition

In our case, the calling syntax is coded as a property of the *MATLAB Algorithm* super class and the algorithm syntax description just contains the name of the corresponding MATLAB function implementing the algorithm functionality.

While useful, neither the description of complex operators' decomposition into more concrete programs (either by specialization –alternatives- or composition - sequences, parallel, loops, etc.-), nor the pre- and post- conditions of use (that state when the operator is applicable and what should hold after its application) will be considered in a first stage. These aspects are left out for future work.

6.2.4.1.2 Data Types

The only property required for every data is its type, which tells the system what kind of values it is allowed to have, thus providing the most important knowledge about it. In a first approximation, the present approach proposes a data type classification into two main groups[128]:

a) Data which are ordinarily referred to as ***images***, i.e. two-dimensional array data and a set of these data (***multi-channel images***).

b) Attribute data of images, such as ***noisiness***, ***brightness***, ***amount of detail***, etc., and parameters of operators, such as ***amount of correction***.

The proposed frame-based representation for data types follows a hierarchical organization based on the following four super classes: ***Image***, corresponding to the first group; and ***String***, ***Number***, and ***Boolean***, for the latter. Each sub-type is then described by the following fields:

- Type name: a symbol describing the type.
- Data type: name of the super class (either ***Image***, ***String***, ***Number*** or ***Boolean***).
- Allowed: allowed values, or limits of the allowed interval (only for numeric types).
- Quotation: reserved for future use.
- Comments (optional): a brief explanation.

The figure shows examples for 'percentage' and 'imageContent' data types descriptions.

```
Type percentage {
  ·Data Type: NumberType
  ·Allowed  : 0,100
  ·Quotation:
  ·Comment  :
 _
}

Type imageContentType {
  ·Data Type: StringType
  ·Allowed  : nature,skin,sky
  ·Quotation:
  ·Comment  :
 _
}
```

Figure 6.5. Example of *types* definition.

Data description may also include information about associated algorithms (this is particularly interesting if processing is ***data-driven***) and display methods [166]. However, for the sake of simplicity, we consider only the latter and, as will be shown, not within the data type description but in the GUI's code.

Finally, we observe that types may also be used as linguistic or symbolic variables. For example, assume that, given an image ***inputImage***, an operator ***measureBrightness*** returns its measured brightness ***imageBrightness***, defined as of type ***percentage***, a previously defined subtype of the type ***Number*** ranging from 0 to 100. Now, we may define the type ***low*** as a new subtype of ***Number***, representing numeric values ranging from 0 to 50. Imagine that, for the actual image, ***imageBrightness*** takes the value 32. Then, the fact '***imageBrightness*** *is (of type)* ***low***' holds true. By omitting the words within brackets, we are using data types as symbolic or linguistic variables.

---

[128] While some domain objects such as ***histogram***, ***line***, ***contour***, etc. may also be used, they highly depend on the specific application and, thus, are not considered within the initial prototype.

6.2.4.2 Knowledge dependent on the contents of an image

This is the knowledge about the expertise in IP, used to process a particular class of images based on their content, characteristics and the final processing purpose. Mostly acquired through observation and based on empirical associations or ***heuristics***, this knowledge consists of cause-and-effect relationships originating from past experiences, such as "*Under a condition X, it is effective to apply image processing Y*". These ***rules of thumb*** are used by the IP expert for algorithm selection, adjustment, composition, execution and control. Indeed, he may not even understand the internal workings of the algorithms, but their symptomatic behavior allows him to associate the inputs with specific outputs. This allows to quickly reach the solution without having to perform a detailed analysis.

6.2.4.2.1 Rules-based procedures

We think that the most appropriate formalism for describing this kind of knowledge are ***IF-THEN*** rules, which cover a wide range of associations such as *condition-action*, *premise-conclusion*, or *antecedent-consequent* [167].

Rules represent conditional knowledge that is quite similar to the way humans express it: as a two-part, condition-action, relationship. A rule can contribute to the problem resolution when the first part, a conditional test (called the ***IF***, the ***condition***, ***premise***, or the ***antecedent***), is satisfied through a true match with known facts. If the rule is *fired*, the second part (called the ***THEN***, the ***action***, the conclusion, or the ***consequent***) is executed, acting on the solution by creating or modifying data.[129]

Rules are the most important element in our system because through them we represent our problem-solving knowledge, not only as an ordered sequence of abstract commands, but by making explicit the relationships that exist between various facts ("results"). They form a network of interconnected facts ("results") in order to define the knowledge within some domain. In this way, the knowledge base is divided into independent and autonomous modules that totally ignore one another. A rule contains expertise to solve some part of the global problem, the solution being built by several rules.

An operator specifies how to perform a task. Each operator is associated with a single task, but a task can be solved by several operators. So one has to tell, for each operator, when it could be used.

In the proposed approach, IP expertise is translated into statements of the form

$$\text{IF X is } A_i \text{ and Y is } B_i \text{ THEN Z is } C_i$$

[129] Notice that, despite looking like conditional statements in procedural programming languages, rules are applied in a totally different manner, making rule-based systems declarative [164], [167].

Basically, a rule links, through an algorithm and under a condition, an ordered set of input results to an ordered set of output results. For example, a rule looks like:

IF *image brightness* is *low*
THEN *corrected image* is *brighten* {*input image*, *$50*}

In this rule, *image brightness* is the antecedent result, *low* is a type, *corrected image* is the computed result, *brighten* is an algorithm's name, *input image* is the input data and *$50* denotes the value of a required numeric parameter (in this case, it is the amount of correction).

In order for the condition '*image brightness* is *low*' to be satisfied, the value of *image brightness* must be among those allowed by the definition of the *low* type. Instead of matching with a type, a literal, numeric or Boolean value may also be used. If that is the case, the quotation specified in each type definition should be used, as is done for the parameter value (*$50*) in the previous example.

Because characteristics of images are essential to direct treatments, first of all the input images must be described using symbolic attributes. To that end, we use *initialization* rules that are always executed thanks to a hard-coded true premise. Then we use rules for the evaluation of images attributes, choice of algorithms and adjustment of their parameters.

RULE R1: ***get black point and white point***
IF TRUE
THEN compute black point and white point as measure histogram extremes of input image

RULE R2: ***stretch histogram***
IF TRUE
THEN compute stretched image as stretch histogram of input image with black point and white point

RULE R3: ***get key***
IF TRUE
THEN compute image key as measure image key of stretched image

RULE R5: ***brighten***
IF image key is dark
THEN compute corrected image as brighten stretched image with amount

RULE R6: ***darken***
IF image key is 'bright'
THEN compute corrected image as darken stretched image with amount

Figure 6.5. Example of an image processing plan for enhancement of tone reproduction

A very good thing is that, when a conclusion is drawn, it is easy to understand how this conclusion was reached. Other advantages of rule-based systems include modularity, uniformity and naturalness. However, their implementation must take special care of issues such as ***infinite chaining***, existence of ***contradictory knowledge***, and ***inconsistent addition of new rules***. These aspects will be discussed in section 6.3.3.1.

6.2.4.3 Facts or Results

Data arguments have fixed values, which are either set (i.e., input data) or computed (i.e., results), while parameters have adjustable values and are always input arguments. All these form the problem-specific data, either provided by the user as initial data about the current problem or progressively derived as it is solved, and are here loosely referred to as ***results***. A result represents data either consumed or produced by an algorithm. This relationship is established through a rule.

Results, which make a statement about some aspect of the problem under analysis and can serve as antecedent or consequent of one or more rules, obtain values during a consultation just as variables obtain values in conventional languages. These values can be determined by any of the following methods: ***1)*** asking the user (through the GUI); ***2)*** using rules to infer the value***; 3)*** obtaining values from an external file or program; ***4)*** using a default value [164].

Apart from its value, a result is internally composed of properties that determine its characteristics. These properties are grouped into three categories:

A. **Required properties**. The most important and only mandatory property of a result is its *type*. This tells the system what kind of values a result is allowed to have. However, note that for uniformity purposes, we decided that every result must have, not only a data object encapsulating its value and type, but also a list of rules that are capable of determining its value.

B. **Internal properties**. These provide the inference engine with the ability to perform the traces that compose the backward-chaining process. These properties are created and maintained internally by the modeling subsystem and the knowledge engineer has no responsibilities on them. The description of these properties is provided as an aid toward understanding how the inference process is performed. These properties are:

   a. ***Production_Rules***: a list of all the backward-chaining rules for this result, i.e. those referencing it within their action part and capable of deriving a value for it. This list, which allows to indirectly obtain the result's ascendants, is used by the modeling subsystem to begin the *backward-chaining* inference process for each goal result.

   b. ***Consuming_Rules***: a list of all the forward-chaining rules for this result, i.e. those referencing it within either their premise ('IF' part) or their action part and unable to set a value for it (what means that the result is used as an input). This allows to indirectly obtain the result's descendents, e.g. in order to invalidate them whenever the result is given a new value.

   c. ***Required***: a Boolean flag indicating whenever the value of a result is being inferred. It is used for detecting infinite backward chaining within the inference process.

   d. ***Version_Stamp***: a version number actualized whenever the result's value is requested. The actual production rule used to derive the result's value

also provides such a version number, which consists of a rule identifier followed by the times that the rule has been fired since the startup of the system.

C. **Optional properties:** these can include a default value and range, help information, specification of the format to be used when presenting or asking the user for the result's value, etc. While some of these properties are rather significant, it is mainly in the way in which they assist the user in interfacing with the system. Thus, since they do not specially contribute to the problem-solving process, they are not considered within our initial prototype.

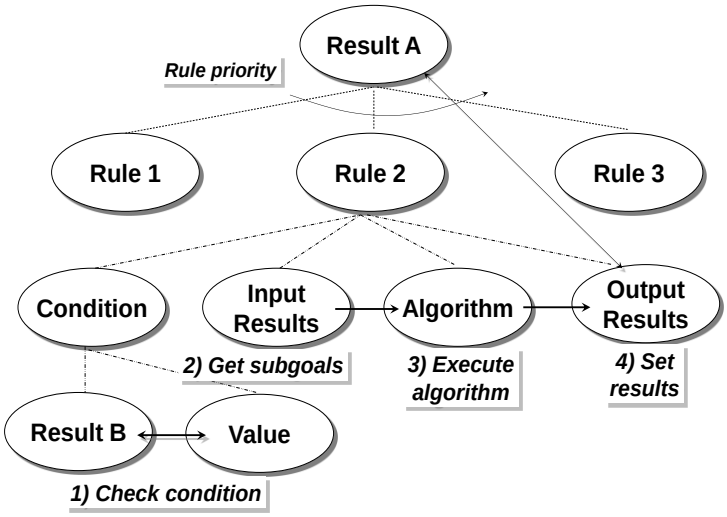


Figure 6.7. Elements interrelation. In order to comput the Result A, we have to *1)* evalutate a set of conditions to properly choose a rule, which in turn *2)* takes a set of previous results, 3) executes an algorithm, which *4)* outputs some new results.

6.2.4.4 Writing a knowledge base

We use plain text files for the specification of algorithms, types and rules. The system supports user-defined syntax, so required information can be given in any arbitrary format, as long as it is first well specified in the file. Such a high flexibility improves system's extensibility and usability, since allows us to use, for instance, natural English-like instead of LISP syntax, which is easier for us to understand, to describe knowledge before it is converted into an internal representation.

This chapter's appendix provides examples of these files that have been used to develop and test the application.

6.2.5 Inference engine: Rule-based reasoning

The main task performed by the modeling subsystem is to decompose a given requirement for image processing into a sequence of IP operators which would lead to an expected image. This process (which involves plan generation, operator selection and parameter adjustment) of deriving a solution to the problem can be viewed simplistically as searching the problem space defined by rules connecting facts, not for any particular piece of data, but rather for a path connecting the initial data to a description of its desired state (i.e., the solved problem). This path, which represents the solution steps of the problem (i.e., the sequence of IP operators), can be found in two main different ways:

6.2.5.1 Data-driven reasoning

In ***data driven reasoning***, the inference engine uses ***forward chaining*** (as it is also called) in order to progress from the initial facts, to intermediate facts, and ultimately to a solution or set of solutions. Rules are typically applied in this way when the number of inputs is limited and/or the number of possible conclusions is large.

6.2.5.2 Goal-directed reasoning

In our case, however, the desired state is unique and given by a specific user defined IP requirement. Thus, since the number of possible final conclusions (called ***goals***) is limited, it is more efficient to follow a ***goal-directed reasoning***, using ***backward chaining***, in order to apply rules only to derive values for goals or for intermediate facts used later to set values of these goals. This process, called ***tracing*** a goal, stops when a goal is either found true or proven to be unsupported because no rule can derive the goal's value. The steps of this process, very close to depth-first search, are depicted in the flow diagram of Figure 6.8.

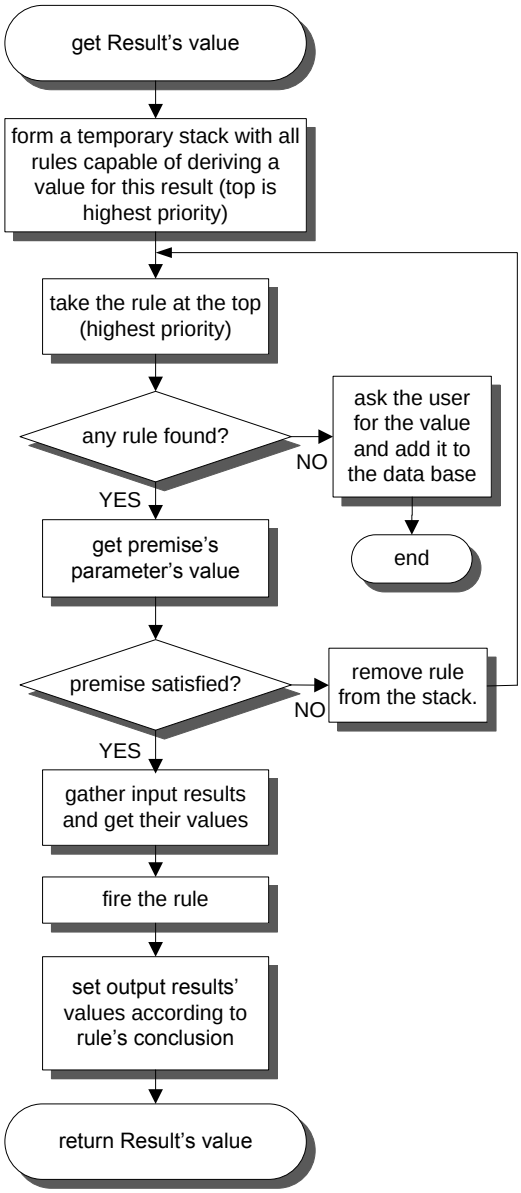


Figure 6.8 Backward reasoning steps

6.3 Implementation details

Each result, either obtained as raw data or derived from other result (i.e. intermediate result) can serve as an antecedent of one or more rules, which draw conclusions that can serve in turn as antecedents of other rules and so forth. Thus, one may visualize the knowledge base as a network of interconnected rules and facts, which are respectively the links and nodes of the resulting hypergraph. When inference is performed, values are derived and set for the various results. This is similar to deriving new facts and placing them within a fact base.

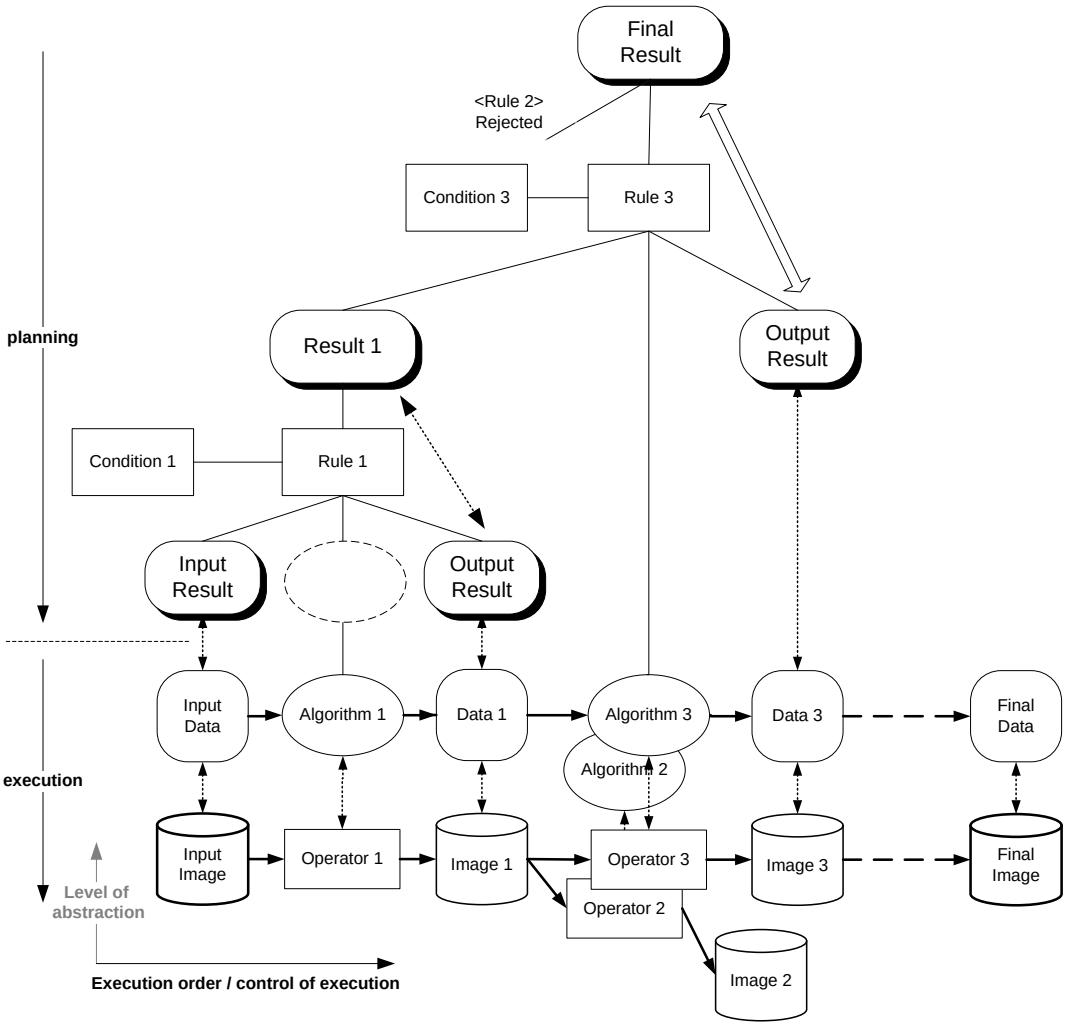


Figure 6.9. Inference structure. A given processing requirement is a main goal to be satisfied, which is divided into several subgoals, each of which corresponds to a specific image processing requirement. These subgoals are decomposed into several subgoals, each of which turns into another set of subgoals, and so on. This subgoaling process continues until the first element of the set of decomposed subgoals corresponds directly to a specific image processing algorithm.

Inference nets like this have been traditionally the architecture used for implementing backward chaining systems. Despite being less flexible and powerful than pattern-matching systems (which have traditionally been used for forward chaining), inference nets are more efficient because all interconnections

can be explicitly stated prior to run-time. This minimizes searching for facts that match premises and simplifies both the implementation of the inference engine and the handling of explanations. In addition, because the explicit interconnections are shown, the conflict resolution problem is reduced to simply maintaining a list of newly matched rules for subsequent firing. Thus, it is easy to decide what resulting actions need to be taken whenever some result is derived.

However, IP problem solving differs from typical planning (among other things) in that the effect of each action, i.e. image processing algorithms, can neither be fully predictable nor describable in a symbolic way. This prevents from determining a complete course of actions before any actual action is performed. The proposed approach is to perform an actual image processing once the subgoaling process reaches a sub goal which correspond directly to an actual algorithm, as is done in [168].

6.3.1 Knowledge Acquisition

A problem with expert systems is writing the rules themselves. Thought processes that are highly rule oriented are easier to write than ones that rely more on creativity or intuition. Another problem is that often experts themselves disagree. Different experts might take different courses of action or go through different thought processes when given the same problem to solve. Thus, there is disagreement in the professional community about the validity of expert systems.

To generate and tune the expert knowledge, images taken from a library were manually enhanced by an IP expert using professional PhotoShop® IP software. In this way, the value of image attributes before and after enhancement was obtained. We also used rules of thumbs that can be found in IP tutorials and do-it-yourself handbooks.

6.3.2 Selection of programing languages

Our goal in developing the application was to offer an integrated simulation and image processing environment where users could implement the algorithms. It was also an attempt to combine the advantages of low-level and high-level languages by borrowing the best from both philosophies. Specifically, we have chosen to base our system on:

- MATLAB [173] as both the algorithm development environment and IP execution sub-system.
- Java as the programming language for the knowledge management and modeling sub-systems, as well as for the GUI.
- JMatLink [174] as connector between the two previous.

Since only standard components and Java are used, the described integrated application should work on all systems that support MATLAB and Java.

6.3.2.1 Development, prototyping and execution of low-level (algorithmic) image processing routines: MATLAB

The two traditional ways to program IP algorithms are through the use of a low-level language, such as C, which offers the advantage of computational speed, an important factor when dealing with images; or a high-level language, such as MATLAB [173], which offers a rich functionality with a large palette of imaging routines, but tends to hide many important aspects of the algorithm.[130]

We decided to use MATLAB for the development, prototyping and execution of image processing algorithms. These are coded as user-defined MATLAB functions, by creating a file of the same name as the new operator that has an ".m" extension and that specifies the number of arguments and returned values using the *function* keyword. Alternatively, an ".m" file can contain a script of commands or standalone program. For faster computation, it is also possible to dynamically link C routines as MATLAB functions through the MEX utility. Instead, we have done our best to use vectorized code, which also makes the program more readable. However, because the MATLAB programming language is imperative, it is limited to algorithm implementation without any "intelligent" capability.

6.3.2.2 Development of knowledge-based system for high-level image processing tasks: Java

We decided to use Java as the programming language for the knowledge management and modeling sub-systems, as well as for the GUI. The decision of using a high-level object-oriented language was based on abstraction, extensibility and maintenance of the sub-systems. The application is divided into several modules, which are constructed by well-defined objects. For portability, the functions depending on the operative system and hardware are grouped into a separate module.

Besides, we have adopted also object-oriented programming to describe the structural knowledge of IP algorithms due to inheritance, abstraction, encapsulation and polymorphism.

---

[130] While many image processing packages have a wide variety of functions, a whole new level of utility and flexibility arises when the image processing functions are built around a programming and/or data analysis environment. Algorithm development environments strive to provide the user with an interface that is much closer to mathematical notation and lexicon than general-purpose programming languages, such as C, C++, or Fortran. The idea is that a user should be able to write the desired computational instructions in a native language that requires relatively little time to master. Algorithms development environments meet this goal by eliminating the compilation step, providing many high-level routines, and guaranteeing portability [17]. This significantly reduces the development time. Additionally, programming environments allow for tailoring image processing techniques to the specific task, developing new algorithms, and interfacing image processing tasks with other scientific data analysis and numerical computational techniques. Also, graphical visualization of the computations should be fully integrated so that the user does not have to leave the environment to observe the output.

The proposed approach uses object-oriented programming to describe the knowledge of the algorithms. Each algorithm, which is an instance of a class of algorithms (see algorithm hierarchy in the figure), consists of the properties mentioned in 6.2.4.1.1 (algorithm name, names and types of arguments, calling syntax and a comment), along with a reference, *engine*, to an object of class *Engine* responsible of the actual algorithm execution and a method, *frame(inputs, outputs)*, invoked to execute the algorithm.[131] These common properties and behavior of every algorithm are described in an abstract superclass *Algorithm*.

```
Algorithm
-MatlabAlgorithm
-JavaAlgorithm
-Operator
   -Equality
   -Greater
   -GreaterOrEq
```

Figure 6.10. Hierarchy of

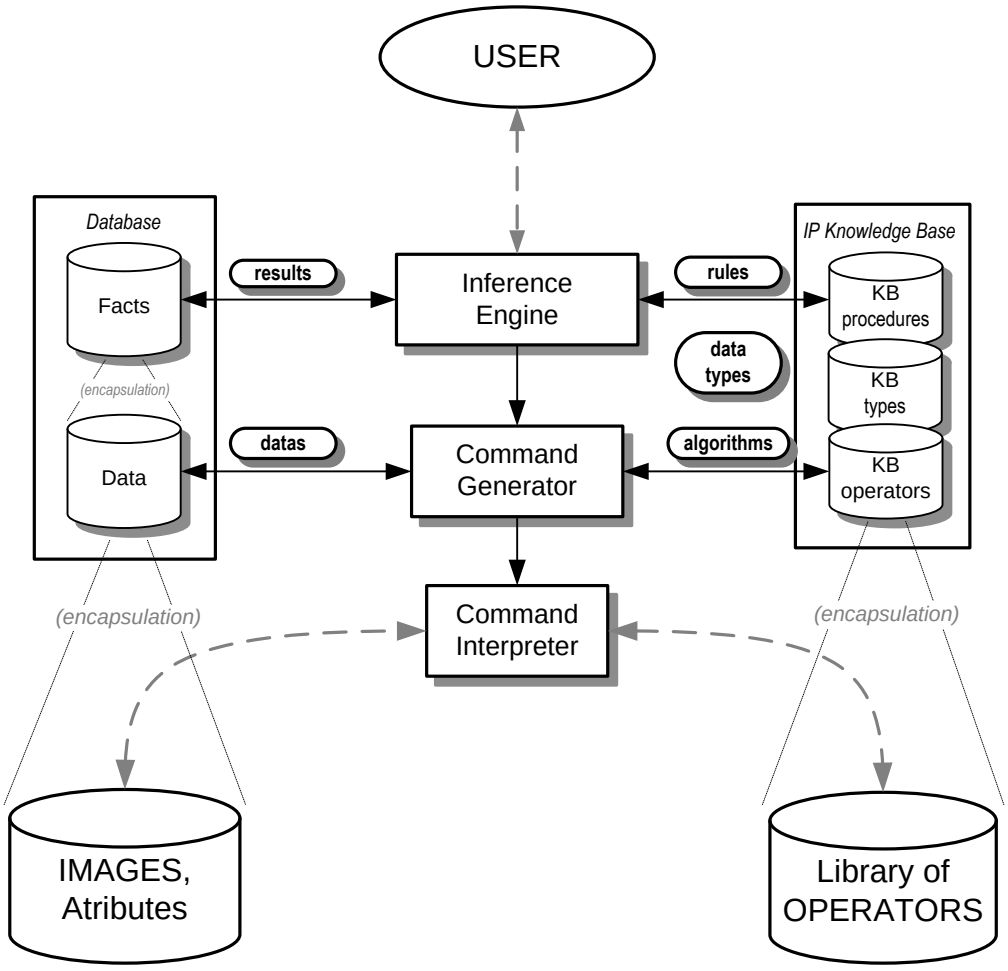


Figure 6.11. System conceptual architecture. While planning is implemented and executed in Java, actual image processing execution is performed in MATLAB. JMatlink serves as interfaces between the two sub-systems.

[131] Construction of the command string is performed by substituting in the calling syntax description with the actual input and output data names.

6.3.3 Disadvantages of rule-based systems

Rule-based systems have, on one hand, three main advantages:

a) ***Modularity***: each rule is a distinct separate unit of knowledge that can be added, modified, or removed independently of the other rules that exist, providing a great flexibility in developing a knowledge base.

b) ***Uniformity***: all knowledge in the system is expressed in exactly the same format, easing the development of the knowledge base.

c) ***Naturalness***: rules are a natural and very common format used by experts to express problem-solving knowledge in many types of domains.

On the other hand, typical problems emerge in the form of ***infinite chaining***, existence of ***contradictory knowledge***, and ***inconsistent addition of new rules***.[132] The implementation of the modeling subsystem has taken special care of these issues. Situations such as these can be extremely difficult to locate and correct as a knowledge base becomes larger and increasingly more complex.

6.3.3.1 Modification of existing Rules and Consistency of the knowledge Base

Contradictions may happen whenever new rules are introduced in the knowledge base, or results are given new values (i.e. data contained in the fact base is modified). Because of the lack of a developed knowledge management subsystem, ensuring the consistency of knowledge base is, by now, up to the knowledge engineer. Nevertheless:

- Conflict resolution is achieved by means of rule priorities. In the proposed system, these are assigned according to the order in which rules are introduced in the knowledge base, being the last added rule the one with highest priority. This simple mechanism provides an easy way of modifying the knowledge base.
- When a result is given a new value, it informs its consuming rules, which in turn invalidate their output results. In this way, forward-chaining is used by a result for propagating its new condition to its descendants, and so forth.

6.3.3.2 Infinite chaining

When adding a new rule, one must take special care that looping does not occur neither within a single rule, nor through several rules. Since such a situation is very hard to be detected through a simple examination of the knowledge base, in the implemented system each result contains a flag named *Required*, which sets whenever its value is being inferred. A loop is found and the corresponding error thrown whenever the system asks a result for its value and its flag is activated (what indicates that the result's value was already being inferred).

---

[132] Many of these problems result from the wrong idea that "*if the system does not work properly, then all one needs to do is adding more rules*" [164].

6.4 Summary

While development of algorithms is usually based on theoretical frameworks, development of image enhancement techniques almost always requires extensive experimental work with large sets of sample images involving algorithm composition, execution, revision and comparison of candidate solutions. Convinced that an explicit knowledge representation together with execution automation capabilities may greatly improve such highly empirical or heuristic tasks, we have covered in this chapter the implementation from scratch of a system based on both generic knowledges about image data types as well as image processing algorithms, and a domain-specific mixture of image enhancement heuristics and procedures taken from image processing and photography communities. Given an image to be processed, a goal to reach (e.g., enhance its quality), and constraints on the result (e.g., naturalness), the system generates and performs the processing.

For the sake of both architectural and processing flexibility, we base our implementation on modularity, separating the different types of knowledge, abstraction and encapsulation. We respectively formalize previous knowledges using structural frame-like declarative descriptions and if-then production rules. Rules are the most important element in our system because through them we represent our problem-solving knowledge, not only as an ordered sequence of abstract commands, but by making explicit the relationships that exist between various facts. Such a system may serve both as experimentation tool for interactive algorithm development and comparison of IP techniques, where algorithms can be applied to images, their attributes may be modified and the interrelation between results studied; and as a IP support tool to enable non-specialists in IP to *i)* developing their own IP applications while limiting their cognitive and skill requirements, and *ii)* exchange knowledge bases for different application purposes (e.g., astronomy, medicine, etc.).

The system is conceived as an initial prototype. During its development, we are mainly concerned with the problems of modeling both the knowledge and reasoning used in IP, being the system's flexibility and generality the only strict requirements we have. Since image quality improvement is a general problem arising in various application domains, we have clearly separated the description of IP procedures from actual algorithm implementation, providing both knowledge models and software tools which are ***general*** (independent of any particular application and of any IP package) and ***flexible***. Following the same principle, we have respectively chosen Java for the expert system and MATLAB for the algorithm implementation, using JMatLink as connector. The several levels of abstraction make possible an incremental development of the system by several specialists, especially when dealing with large knowledge bases.

Last, but not least, we are not only interested in solving IP problems, we also want to understand the reasoning of the system, in order to improve its results

and behavior, and, in the long term, to allow explaining and teaching Image Processing to non-expert users. We hope that this chapter will attract the attention of many readers to the application of knowledge-based techniques for developing new IP systems. We are optimistic that the provided ideas and developed tools will be useful in adding value to the way IP is research and applied by non-experts.

Future work

Since we are dealing with images, our knowledge-based system should be provided with a multi-window graphical user interface for result visualization as well as interactive plan creation and execution. Users could adopt either a bottom-up building process (grouping subtasks together), or a top-down one (decomposing into sub-tasks), or even a mixed one. This interface could establish a dialogue with the user, in order to have him clarify the specifications of its request, and also include graphical tools to show details directly on images. During execution, users could visualize intermediate images that are input or output of any task of the plan, or have access to data about tasks and solving methods, through the graphical figure of the plan as a tree of tasks.

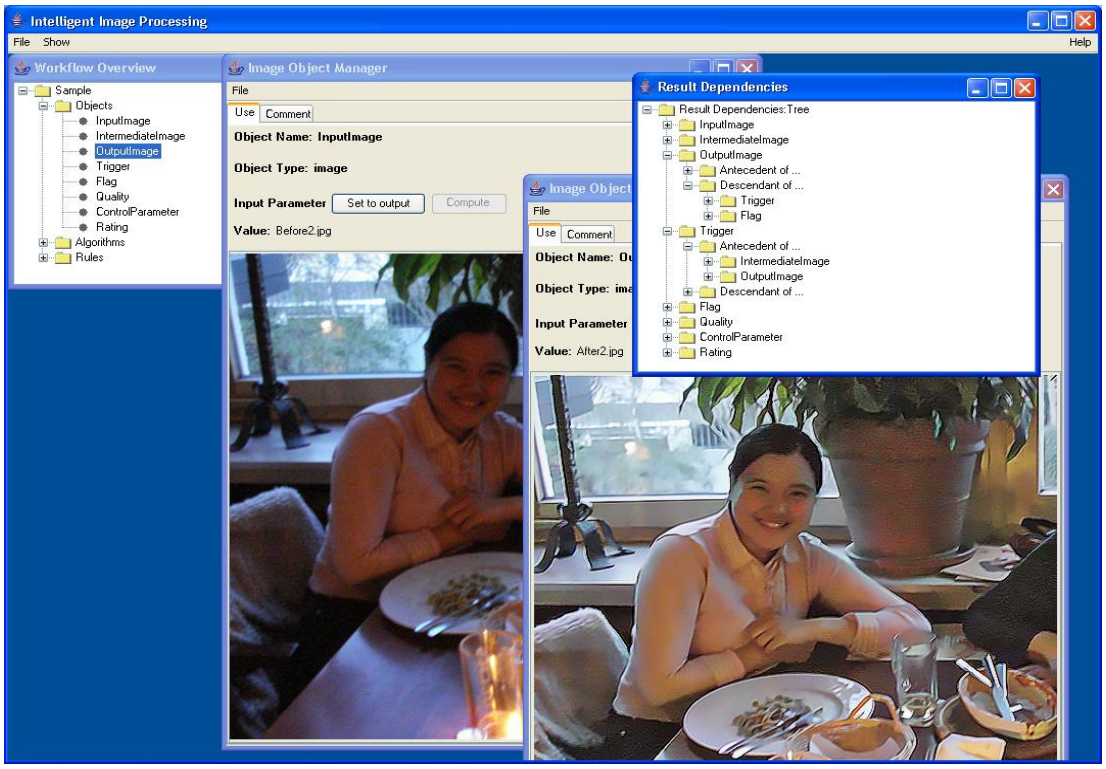


Figure 6.12. Multi-window application snapshots. This is how our application prototipe would look like when using a multi-window GUI. The user has simultaneous access to the *workflow* or plan overview, input, intermediate and output images, as well as generated dependencies.

Areas of future research include *specification by example* and fricher knowledge modeling, for instance

- Fuzzy rules, as in [161].
- Richer algorithm behavior description.
- Graph representation of workflow instead of production rules

6.5 Appendix

6.5.1 Algorithms

```
/*
Note: First line is used to know comment indicator string. Second and third
lines indicate comment block start and end. Do not write there.

*****************************************************************************
                        - MATLAB ENGINE ALGORITHMS FILE -
*****************************************************************************

ALGORITHM DESCRIPTION FORMAT:

 -Algorithm information can be given in any arbitrary format, as long as it
 is first well specified. The format of a rule information is identified by
 the reserved word

         ·INFO

 -The next are the field names of an algorithm description. You are allowed to
use them in any order to describe the algorithm format used in this file.
Be sure that all of them are used in the format description.

          ·AlgorithmName (used by programm as identifier)
          ·ClassName
          ·MATLABName
          ·InputNames     (for user understability, not used by programm)
          ·OutputNames    (for user understability, not used by programm)
          ·InputTypes
          ·OutputTypes
          ·Comment        (for user understability, not used by programm)

If no specific java wrapper class is implemented, 'ClassName' field should have
'workflow.engine.MATLAB.MATLABAlgorithm' value. Otherwise, specify the fully
qualified name of wrapper java class.

 -You can use pseudo field names. Their content will be read but not used.

         (OPTIONAL->)

 -Head and Tail to delimit each algorithm should also be provided, as well as a
Delimiter for the fields and a separator for multiple field contents.
For this purpose, the next reserved words are used:

          ·HEAD
          ·TAIL
          ·DEL.
          ·SEP.


 Example of Format Description:

         FORMAT DESCRIPTION START
         /Algorithm/ 'AlgorithmName' {
         ·Wrapper Name: 'ClassName'
         ·MATLAB  Name: 'MATLABName'
         ·Input  Names: 'InputNames'
         ·Input  Types: 'InputTypes'
         ·Output Names: 'OutputNames'
         ·Output Types: 'OutputTypes'
         ·Comment: 'Comment'
         /}/
         DELIM.:'
         SEPAR.:,
         FORMAT DESCRIPTION END
```

So an algorithm description looks like:

```
Algorithm 'AlgorithmName' {
·Wrapper     : 'WrapperClass'
·MATLAB  Name: 'MATLABName'
·Input  Names: 'InputNames'
·Input  Types: 'InputTypes'
·Output Names: 'OutputNames'
·Output Types: 'OutputTypes'
·Comment: 'Comment'
}
```

```
MATLAB is a registered trademark of The Mathworks Incorporated
*/

//FORMAT SYNTAX AND DESCRIPTION:

FORMAT SYNTAX DELIM:"
FORMAT SYNTAX SEPAR:%
FORMAT SYNTAX START:
/"HEAD"/"INFO"/"TAIL"/
DELIM.:"DEL."
SEPAR.:"SEP."
FORMAT SYNTAX END

FORMAT DESCRIPTION START
/Algorithm/ 'AlgorithmName' {
·Wrapper     : 'ClassName'
·MATLAB  Name: 'MATLABName'
·Input  Names: 'InputNames'
·Input  Types: 'InputTypes'
·Output Names: 'OutputNames'
·Output Types: 'OutputTypes'
·Comment:
'Comment'
/}/
DELIM.:'
SEPAR.:,
FORMAT DESCRIPTION END


//ALGORITHM DESCRIPTIONS START HERE:

//ALGORITHMS USED BY CONDITION RULES

Algorithm equals {
·Wrapper     : Equality
·MATLAB  Name: -
·Input  Names: x,y,inverter
·Input  Types: String,String,Boolean
·Output Names: comparison
·Output Types: Boolean
·Comment:
Returns Boolean.TRUE if x == y.
}

Algorithm greater than {
·Wrapper     : Greater
·MATLAB  Name: -
·Input  Names: x,y,inverter
·Input  Types: Number,Number,Boolean
·Output Names: comparison
·Output Types: Boolean
·Comment:
Returns Boolean.TRUE if x > y.
}

Algorithm greater than or equal to {
·Wrapper     : GreaterOrEq
·MATLAB  Name: -
·Input  Names: x,y,inverter
·Input  Types: Number,Number,Boolean
·Output Names: comparison
·Output Types: Boolean
·Comment:
Returns Boolean.TRUE if x > y.
}


// ALGORITHMS USED BY ENGINE

Algorithm SAVE {
·Wrapper     : MATLABAlgorithm
·MATLAB  Name: saveToFile
·Input Names:
variable,fileName,format
·Input  Types: Image,String,String
·Output Names:
·Output Types:
·Comment:
TODO change input type 'Image' to
'Generic'
}

Algorithm LOAD {
·Wrapper     : MATLABAlgorithm
·MATLAB  Name: loadFromFile
·Input  Names: fileName
·Input  Types: String
·Output Names: variable
·Output Types: Image
·Comment:
TODO change output type 'Image' to
'Generic'
}

// Next algorithms use the generic
java wrapper class 'MATLABAlgorithm'

Algorithm brighten {
·Wrapper     : MATLABAlgorithm
·MATLAB  Name: brighten
·Input  Names: image,amount
·Input  Types: Image,percentage
·Output Names: brightened image
·Output Types: Image
·Comment:
}
```

```
Algorithm darken {
·Wrapper     : MATLABAlgorithm
·MATLAB  Name: darken
·Input  Names: image,amount
·Input  Types: Image,percentage
·Output Names: darkened image
·Output Types: Image
·Comment:
}

Algorithm stretch histogram of {
·Wrapper     : MATLABAlgorithm
·MATLAB  Name: stretchHist
·Input  Names:
image,blackPoint,whitePoint
·Input  Types:
Image,pixelLevel,pixelLevel
·Output Names: stretched image
·Output Types: Image
·Comment:
}

Algorithm measure histogram extremes
of {
·Wrapper     : MATLABAlgorithm
·MATLAB  Name: measHistXtr
·Input  Names: image,percentile
·Input  Types: Image,percentage
·Output Names: blackPoint,whitePoint
·Output Types: pixelLevel,pixelLevel
·Comment:
}

Algorithm measure image content of {
·Wrapper     : MATLABAlgorithm
·MATLAB  Name: measImgContent
·Input  Names: image
·Input  Types: Image
·Output Names: content
·Output Types: imageContentType
·Comment:
}

Algorithm measure image key of {
·Wrapper     : MATLABAlgorithm
·MATLAB  Name: measImgKey
·Input  Names: image
·Input  Types: Image
·Output Names: key
·Output Types: imageKeyType
·Comment:
}
```

6.5.2 Data Types

```
/*
Note: First line is used to know comment indicator string. Second and third
lines indicates comment block start and end. Do not write there.

*************************************************************************
                                - TYPES FILE -
*************************************************************************

TYPE DESCRIPTION FORMAT:

 -Type information can be given in any arbitrary format, as long as it is first
well specified. The format of a type information is identified by the reserved
word

      ·INFO

 -The next are the field names of a type description. You are allowed to use
them in any order to describe the type format used in this file.
 Be sure that all of them are used in the format description.

      ·TypeName
      ·ClassName
      ·Allowed
      ·Quotation
      ·Comment

 -'ClassName' field can take the predefined values:

      ·workflow.model.types.DataType   (for generic type)
      ·workflow.model.types.NumberType (for numbers)
      ·workflow.model.types.StringType (for string)
      ·workflow.model.types.ImageType  (for images)

 or any other that corresponds with the name of the java class that
 implements it.

 -'Allowed' field can take the values:

      ·any           (to allow every possible value)
      ·min,max       (to allow every numeric value contained in [min,max])
      ·str1,str2,... (to allow every string value contained in the provided set)


 -You can use pseudo field names. Their content will be read but not used.

      ·(OPTIONAL->)

 -Head and Tail to delimit each type should also be provided, as well as
 a Delimiter for the fields and a Separator for multiple field contents.
 For this purpose, the next reserved words are used:

      ·HEAD
      ·TAIL
      ·DEL.
      ·SEP.



Example of Format Description:                    'Comment'
                                                  /}/
         FORMAT DESCRIPTION START                 DELIM.:'
         /Type/ 'TypeName' {                      SEPAR.:,
         ·Data Type: 'ClassName'                  FORMAT DESCRIPTION END
         ·Allowed  : 'Allowed'
         ·Quotation: 'Quotation'
         ·Comment  :
```

```
So a type description looks like:

        Type 'TypeName' {
        ·Data Type: 'ClassName'
        ·Allowed  : 'Allowed'
        ·Quotation: 'Quotation'
        ·Comment  : 'Comment'
        }

*/

//FORMAT SYNTAX AND DESCRIPTION:

FORMAT SYNTAX DELIM:"
FORMAT SYNTAX SEPAR:%
FORMAT SYNTAX START:
/"HEAD"/"INFO"/"TAIL"/
DELIM.:"DEL."
SEPAR.:"SEP."
FORMAT SYNTAX END

FORMAT DESCRIPTION START
/Type/ 'TypeName' {
·Data Type: 'ClassName'
·Allowed  : 'Allowed'
·Quotation: 'Quotation'
·Comment  :
'Comment'
/}/
DELIM.:'
SEPAR.:,
FORMAT DESCRIPTION END


//TYPE DESCRIPTIONS START HERE:

//APPLICATION TYPES. THESE
DESCRIPTIONS SHOULD NOT BE CHANGED.

Type Generic {
·Data Type: DataType
·Allowed  : any
·Quotation:
·Comment  :
-
the     name     is     defined     by
'DataType.GenericType_ID'     static
constant
}

Type Image {
·Data Type: ImageType
·Allowed  : any
·Quotation: "
·Comment  :
-
}

Type String {
·Data Type: StringType
·Allowed  : any
·Quotation: '
·Comment  :
-
}

Type Number {
·Data Type: NumberType
·Allowed  : any
·Quotation: $
·Comment  :
-
}

Type Boolean {
·Data Type: BooleanType
·Allowed  : any
·Quotation:
·Comment  :
-
}


// USER DEFINED TYPES:

Type percentage {
·Data Type: NumberType
·Allowed  : 0,100
·Quotation:
·Comment  :
-
}

Type pixelLevel {
·Data Type: NumberType
·Allowed  : 0,1
·Quotation:
·Comment  :
-
}

Type imageContentType {
·Data Type: StringType
·Allowed  : nature,skin,sky
·Quotation:
·Comment  :
-
}

Type imageKeyType {
·Data Type: StringType
·Allowed                                :
veryDark,dark,bright,veryBright
·Quotation:
·Comment  :
-
}

Type dark {
·Data Type: StringType
·Allowed  : veryDark,dark
·Quotation:
·Comment  :
-
}

Type bright {
·Data Type: StringType
·Allowed  : bright,veryBright
·Quotation:
·Comment  :
-
}
```

6.5.3 Rules

```
/*
Note: First line is used to know comment indicator string. Second and
third lines indicate comment block start and end. Do not write there.

**********************************************************************
*                           - RULES FILE -                           *
**********************************************************************

RULE DESCRIPTION FORMAT:

 -Rule information can be given in any arbitrary format, as long as it is
 first well specified. The format of a rule information is identified by
 the reserved word

     ·INFO

 -The next are the field names of a rule. You are allowed to use them in
  any order to describe the rule information format used in this file.
  Be sure that all of them are used in the format description.

     ·RuleName
     ·AlgorithmName
     ·InputNames
     ·OutputNames
     ·Condition
     ·Comment

 -You can use pseudo field names. Their content will be read but not used.
 For example:

     ·(OPTIONAL->)

 -Head and Tail to delimit each rule should also be provided, as well as a
 Delimiter for the fields and a Separator for multiple field contents.
 For this purpose, the next reserved words are used:

     ·HEAD
     ·TAIL
     ·DEL.
     ·SEP.


 Example of Format Description:

     FORMAT DESCRIPTION START
     ->HEAD:Rule:
     ->INFO:'RuleName'  ['OutputNames']  =  'AlgorithmName'('InputNames')'o',  if
'Condition';
     'Comment'
     ->TAIL:.
     ->DELI:'
     ->SEPA:,
     FORMAT DESCRIPTION END


 So a rule description looks like:

         Rule: First rule [R1,R2] = Mswap(Ra,Rb), if Ra is 5;
         (This is an example rule, and now you are reading
         the comment field.)
         end


         Note that some fields can be left blanc:
```

```
        Rule:  [R1,R2] = Mswap(Ra,Rb), if Ra is 5
        end
*/

//FORMAT SYNTAX AND DESCRIPTION:

  FORMAT SYNTAX DELIM:"
  FORMAT SYNTAX SEPAR:%
  FORMAT SYNTAX START:
  ->HEAD:"HEAD"
  ->INFO:"INFO"
  ->TAIL:"TAIL"
  ->DELI:"DEL."
  ->SEPA:"SEP."
  FORMAT SYNTAX END

  FORMAT DESCRIPTION START
  ->HEAD:RULE
  ->INFO: 'RuleNumber': 'RuleName'
        IF      'Condition'
        THEN    compute  'OutputNames'  'AlgorithmName'  'InputNames'
  'Comment'
  ->TAIL:END
  ->DELI:'
  ->SEPA:, | and | with
  FORMAT DESCRIPTION END

//RULE DESCRIPTIONS START HERE:
//For structure clarity purposes, rule comments have been omited.
```

RULE R1: ***get black point and white point***

```
   IF    TRUE
   THEN  compute black point and white point  as measure histogram extremes of
input image with $1$
END
```

RULE R2: ***stretch histogram***

```
   IF    TRUE
   THEN  compute  stretched image  as stretch histogram of  input image with
black point and white point
END
```

RULE R3: ***get key***

```
   IF    TRUE
   THEN  compute  image key  as measure image key of  stretched image
END
```

RULE R4: ***get content***

```
   IF    TRUE
   THEN  compute  image content  as measure image content of  stretched image
END
```

RULE R5: ***brighten***

```
   IF    image key  is  dark
   THEN  compute  corrected image  as brighten  stretched image with amount
END
```

RULE R6: ***darken***

```
   IF    image key  is  'bright'
   THEN  compute  corrected image  as darken  stretched image with amount
END
```

RULE R7: ***skin case***

```
   IF    image content  is  'skin'
   THEN  compute  corrected image  as  stretched image
END
```

RULE R8: ***skin case***

```
   IF    black point  is greater than  $1$
   THEN  compute  corrected image  as  input image
END
```

REFERENCES

[158] CLÉMENT, Véronique; THONNAT, Monique. *Integration of image processing procedures: OCAPI, a knowledge-based approach*. 1990. INRIA.

[159] CLOUARD, Régis, et al. Borg: A knowledge-based system for automatic generation of image processing programs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, vol. 21, no 2, p. 128-144.

[160] CLOUARD, Régis, et al. Resolution of image processing problems by dynamic planning within the framework of the Blackboard model. En *SPIE Int. Symposium: Intelligent Robot and Computer Vision XII: Algorithms and Techniques*. 1993. p. 419-429.

[161] EL-OSERY, A.; JAMSHIDI, Mohammad. Analog and digital image enhancement using fuzzy expert systems. *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, 2002, vol. 7, no 2, p. 97-106

[162] FICET-CAUCHARD, Valérie; PORQUET, Christine; REVENU, Marinette. An interactive case-based reasoning system for the development of image processing applications. *Advances in Case-Based Reasoning*, 1998, p. 437-447.

[163] GIARRATANO, J.; RILEY, G. Expert systems: Principles and programming, 1998. *Boston, MA, PWS-Kent Publishing*.

[164] GONZALEZ, Avelino J.; DANKEL, Douglas D. *The engineering of knowledge-based systems: theory and practice*. Prentice-Hall, Inc., 1993.

[165] MATSUYAMA, Takashi. Expert systems for image processing-knowledge-based composition of image analysis processes. En *Pattern Recognition, 1988., 9th International Conference on*. IEEE, 1988. p. 125-133.

[166] MOISAN, Sabine; VINCENT, Régis; THONNAT, Monique. *Program supervision: from knowledge modeling to dedicated engines*. 1997. Tesis Doctoral. INRIA.

[167] RUDOLPH, George. Some guidelines for deciding whether to use a rules engine. *Retrieved June*, 2003, vol. 13, p. 2005.

[168] TANAKA, Toshikazu; SUEDA, Naomichi. Knowledge acquisition in image processing expert system'EXPLAIN'. En *Artificial Intelligence for Industrial Applications, 1988. IEEE AI'88., Proceedings of the International Workshop on*. IEEE, 1988. p. 267-272.

[169] TANIGUCHI, R.; AMAMIYA, M.; KAWAGUCHI, E. Knowledge-based image processing system: IPSSENS-II. En *Image Processing and its Applications, 1989., Third International Conference on*. IET, 1989. p. 462-466.

[170] THONNAT, Monique; MOISAN, Sabine; CRUBÉZY, Monica. Experience in integrating image processing programs. *Computer Vision Systems*, 1999, p. 200-215.

[171] ZAVIDOVIQUE, Bertrand; SERFATY, Veronique; FORTUNEL, Christian. Mechanism to capture and communicate image-processing expertise. *IEEE Software*, 1991, vol. 8, no 6, p. 37-50.

[172] Khoros Pro 2001 Integrated Development Environment, Khoral Inc., Albuquerque, NM [Online]. Available: http://www.khoral.com/

[173] The MathWorks Inc. (MATLAB), Natwick, MA {Online}. Available: http://www.mathworks.com/

[174] Müller, S.: 1999. JMatLink library. http://homepage.ruhr-uni-bochum.de/Stefan.Mueller/JMatLink/.

[175] W. Rasband, ImageJ, National Institutes of Health, Bethesda, MD. Available: http://rsb.info.nih.gov/ij/

Chapter 7

CONCLUSION

SUMMARY

Improving visual information is a primary requirement for almost all vision and image processing tasks, for which a huge amount of image processing algorithms has been developed. On one hand, image restoration is commonly seen as a set of ill-posed inverse problems, for which sophisticated mathematical theories have been proposed. On the other, image enhancement, regarded as a much more subjective issue, has been largely dominated by heuristics. As yet, even very low-level visual processing remains a challenging problem.

In this context, the presented thesis deals with improving perceived quality of digital photographs as an automatic process. Image quality, traditionally concerned with the measurement of distortions introduced by processes such as coding and compression, is here understood as the degree to which the perceptions of the scene observer and the reproduction observer match each other. Furthermore, it is considered to be mainly limited by the inherent degradation and lack of perceptual constancy of the capture-reproduction process. However, rather than contributing with new psychometrical scaling to the wealth of emerging appearance models, the problem is posed in terms of providing qualitative research and appropriate tools for early processing, mainly concerned with extracting intrinsic properties from the image. Special effort is done to remark that skipping perceptual constancies we usually take for granted at this level can turn effortless tasks into very difficult puzzles.

Inspired by the early human visual system, restoration and enhancement are first viewed as an estimation problem: *What are the physical properties of the scene most likely to explain the sensory input?* This unifies both areas of image processing and places them on common ground with research fields such as visual perception, computer vision or information theory. In a second step, the captured image is then transformed to better resemble the original scene in accordance to its inferred physical properties. Within the proposed framework, the minimum set of separable appearance attributes which influence quality in most situations is well known to be composed of independent dimensions such as graininess, brightness and sharpness. For the first one, state of the art edge-preserving smoothing algorithms based on robust statistics in spatial domain are reviewed.

For tone reproduction and detail enhancement, classical low-level vision theories based on adaptation and local interactions at a physiological level are

gathered to provide an overview of centre-surround processes as background for the following development of simple algorithms. Those simple in formulation, ubiquitous and efficient are chosen, slightly modified and eventually implemented in MATLAB®.

Although similar algorithms and more powerful ones have been widely integrated in many available program libraries, no support has been provided to the user without enough expertise for digital image processing to solve practical problems such the one here considered. Complex image processing tasks require selecting the appropriate algorithms and setting the correct parameters values according to the contents and characteristics of the given image and, therefore, are often difficult to fine-tune. Moreover, extensive experimental work is required to develop image enhancement techniques, in which algorithm composition, execution and control are highly based on empirical or heuristic knowledge. As a result, routine application, when feasible, is rather limited.

In order to overcome these limitations and enable end-users to accomplish complex image processing tasks while at the same time limiting their cognitive and skill requirements, a system is devised in which expert's knowledge is explicitly stated in the form of rules. Developed with classical knowledge-based techniques and finally implemented in Java, the proposed system allows easy adaptation to specific tasks by exchanging knowledge bases for different areas like computer vision, remote sensing or medical image analysis.

Finally, evaluation of its performance with a survey of observer's opinions concludes with positive results. In addition, presented concepts and developed tools are general enough to cover a wide range of applications where producing digital images with low noise, good tone reproduction and visible detail is a strong requirement. However, these are just first steps in a new direction in understanding of what perceived image quality is and how it can be automatically improved. Future work necessarily includes evolution of rule representation schemes to more versatile ones as well as the addition of learning capabilities.

We hope that readers will enjoy reading this thesis work as much as we have enjoyed its research and writing. We hope also they find materials provided in it timely, stimulating, useful and relevant to their work and studies in the field. We insist that the framework proposed in this thesis is not intended as a rigid set of boxes, but to provide a common framework and raise issues. Although it offers some practical insights, it is intended more as an in-breadth overview and starting point.

7.1 Contributions

This thesis contributes with both theoretical principles and software implementation of digital image processing concepts. Techniques, tools and ideas developed are not completely new, but an extension of classical and state-of-the-art techniques, contributing to progress towards an efficient unification of image processing techniques and models of human visual perception. By putting all them together and analyzing their interrelation, a solid basis of theoretical framework is prvided, which will support subsequent work for the final goal of developing systems that able to automatically improve the quality of images (and even learn it).

Mayor contributions of this thesis include the following:

- Formulation of the whole image quality improvement as an estimation process. Addition of the lack of perceptual constancy from which capturing and displaying devices suffer to classical degradation models, posing restoration and enhancement as two faces of the same estimation problem: *What are the physical properties of the scene most likely to explain the sensory input?* This unifies both areas of image processing and places them on common ground with research fields such as visual perception, computer vision or information theory, for which a state-of-the-art review is done.

- Establishment of the relation between classical regularization, Bayesian approach and Robust statistics approach to edge-preserving image smoothing.

- New approach to no-reference quality metrics and quality improvement, extending very state-of-the-art approaches based on information theory.

- Development of a rule-based system for aided image processing, conceived as a knowledge-based system for prototyping and automation of complex image processing task within the context of image quality improvement. Knowledge about algorithms and data types is explicitly stated using frame-like declarative structures, while IP expertise is explicitly stated in the form of rules.

Presented concepts and developed tools are general enough to cover a wide range of applications where producing digital images with low noise, good tone reproduction and visible detail is a strong requirement. However, these are just first steps in a new direction in understanding of what perceived image quality is and how it can be automatically improved. Future work necessarily includes evolution of rule representation schemes to more versatile ones as well as the addition of learning capabilities. Smart depiction and visual communication raise a wealth of exciting issues for future work, and interdisciplinary approaches like this are a key to success.

7.2 Examples

Below we show some original photographs that have very low quality because are both *unnatural*, as they fail to peoperly resemble the visual appearance of the scene, and *useless*, as the details carrying relevant information are hardly visible. Together we show the improved-quality images that result from automated application of the noise reduction and tone reproduction algorithms, properly tuned to each image according to a set of heuristic rules provided by an image processing expert.

a) original high contrast image

b) improved image

c) original low-light image

d) improved image

Figure 7.1. Comparison of low quality input images and high quality output images, that result from automated knowledge-based application of the algorithms and techniques described in this thesis.

7.3 Future Work

The limited space of this thesis has allowed us to introduce the basic problems, ideas and exemplar approaches to image quality assessment and improvement. The general methods discussed in this thesis are certainly extendable to many other areas, where improving the performance of described methods is also possible if they are to be applied to specific applications. First, the distortion types are usually constrained and predictable for given application environments, and the measures that can directly quantify these application-specific distortions may provide useful indications of image quality. Second, specific applications are typically associated with specific visual tasks. For example, the ability to visually detect certain objects would be a very important factor for assessing the quality of medical images.

Despite of the high interest in scientific and medical purposes, image quality in terms of visual information capacity as given by Shannon's formula has almost not been studied before. The idea is here presented somewhat informally. Future work would include a more rigorous theoretical formulation.

Multiscale image representation has almost not been covered, which is the main critique to the image processing part. The most sensible election is wavelet decomposition, which has shown to be very powerful. Future work should, with no doubt, explore methods and techniques in the wavelet transform domain.

We observed that keeping the noise estimator as a separate module, which may be replaced with better technique if one becomes available, may however yield to suboptimal solutions. Ideally, the processes of noise estimation and denoising should be intimately merged in one. Moreover, the Bayesian framework provides a formal way for choosing appropriate tonal kernels for the data and smoothness terms, restricting the parameter space depending on the noise. Studying other types of noise and the properties of the signal to recover will lead to different criteria for selecting the penalisers.

Tone reproduction would really benefit from an update based on very recent research incorporating newest models for the recovery of intrinsic images from a single image. It would be also convenient to perform an extensive validation of tone reproduction operators, preferably through psychophysical comparison.

Finally, our knowledge-based system should be provided with a multi-window graphical user interface for result visualization as well as interactive plan creation and execution. This interface would establish a dialogue with the user, in order to have him clarify the specifications of its request, and also include graphical tools to show details directly on images. Areas of future research include specification by example and fricher knowledge modeling.